2018-04-20

Maelstrom Represses Canonical RNA Polymerase II Transcription in Drosophila Dual-Strand piRNA Clusters

Timothy H. Chang
University of Massachusetts Medical School

MAELSTROM REPRESSES CANONICAL RNA POLYMERASE II TRANSCRIPTION IN *DROSOPHILA* DUAL-STRAND piRNA CLUSTERS

A Dissertation Presented

By

TIMOTHY HAN CHANG

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

April 20, 2018

M.D./Ph.D. Program in Biomedical Sciences

MAELSTROM REPRESSES CANONICAL RNA POLYMERASE II TRANSCRIPTION IN *DROSOPHILA* DUAL-STRAND piRNA CLUSTERS

A Dissertation Presented

By

TIMOTHY HAN CHANG

This work was undertaken in the Graduate School of Biomedical Sciences
M.D./Ph.D. Program in Biomedical Sciences

Under the mentorship of

_____
Phillip D. Zamore, Ph.D., Thesis Advisor

_____
Craig L. Peterson, Ph.D., Member of Committee

_____
Oliver J. Rando, M.D./Ph.D., Member of Committee

_____
Zhiping Weng, Ph.D., Member of Committee

_____
Katalin Fejes Tóth, M.D./Ph.D., External Member of Committee

_____
William E. Theurkauf, Ph.D., Chair of Committee

April, 20 2018

DEDICATION

I dedicate this thesis to my loving parents;

David and Judy

my always hungry brothers;

Paul and Samuel

my inspiring grandparents;

Gong Gong, Qing Niang, Wai Gong, and Wai Po

and last but not least, my wonderful wife and son;

Jess and Kai

ACKNOWLEDGEMENTS

Since I was a child, my parents have encouraged me to pursue my interest in observing the natural world. I still clearly remember my mom biking me, bottle in hand, to the nearby park to catch bugs—which I only later realized she found gross. It turns out this was only the least of her sacrifices. Although I have only been a parent for less than a year, I cannot begin to comprehend all that my parents have given up for me and my brothers. For example, during every family vacation my dad would spend time in the hotel room to work or every day my mom would sit in rush hour traffic to try to get home early to take me and my brothers to various lessons and activities. I am finally starting to realize that despite my protests growing up, all of my parents' advice and admonitions are turning out to be true. While I never would have thought this, the phrase "I am turning into my parents" is both the highest compliment I can imagine and something I truly hope that is happening to me.

While I cannot thank my parents enough, even my parents will readily admit that they are also the products of their own parents' love and hard work. There is no way I would be where I am today without my grandparents. Today's world can seem frightening and unstable, but I cannot imagine the world my grandparents lived through: first fleeing their homes during war and again due to a hostile government, they left everything behind. Starting from nothing, they raised and provided for their families and eventually emigrated to the United States when they realized that coming here would offer the best opportunities for their families. Since I was a child, whether it was teaching me to ride a bike, telling me stories and jokes, and even helping to provide for my education, my grandparents have always supported me.

People often say that being a middle child is tough, but in my case, I don't think I would have it any other way. My brothers have always been my best friends and I look up to them in every aspect. There aren't many people that are more genuinely curious

about anything and everything than my older brother, Paul. As a kid, this was very annoying: you could never get his attention when he was reading and when he wasn't reading, he would constantly bug you with questions. Furthermore, he would strike up conversations with everyone. I vividly remember burying my face in shame when he started asking a taxi driver about where he was born. Somehow Paul knew all about the political nuances in this taxi-driver's home country and ten minutes into the ride they were chatting like old friends. My younger brother, Samuel, is the hardest worker and quite possibly the smartest person I know. Sam felt like he had to live up to both his older brothers' reputations in high school (as small as they may be) and rather than just pick one or two things Paul and I did, he did everything we did but did them better. Not only was Sam editor-in-chief of the school newspaper, he was also captain of the swim and water polo teams, and valedictorian too. Way to make us look bad, Sam. Through every major and minor step in my life, my brothers have always been there to support me and provide me with advice and perspective.

Throughout graduate school, Phil has been an amazing mentor and role model. Thinking back to the start of graduate school and how I was as an early scientist, I am deeply appreciative that Phil accepted me into his lab and took on the challenge to shape me into the scientist I am today. While I still have a ways to go, there are many things that I will forever take with me, among them: "ticks inside," the importance of good controls, and "if it doesn't have numbers, it's not science." This thesis would also not be possible without the guidance of my TRAC: Bill, Zhiping, Ollie, and Craig. Moreover, Tiff and our lab mom, Gwen, are responsible for the smooth day-to-day operations of the lab. Without Gwen, no experiments would ever get done and without Tiff, nothing would ever get scheduled.

I would also like to thank all the other members of the Zamore Lab for being wonderful friends and fostering a great scientific community. They made coming to lab everyday fun and exciting. In particular, there is no way I could be where am I without

my other mentor and brilliant scientist, Chengjian. He first took me on as a rotation student and eventually became like a big brother; teaching me to become a scientist and also looking after me at the same time. Without Cindy and Alicia maintaining the fly room much of this work would not be possible. Cindy also helped with fly dissections and our wonderful conversations in the fly room always made fly work enjoyable. Jen has also been a mentor since the beginning with Chengjian. The list of techniques she has taught me is too much to list, but it's the things outside of lab that I will forever remember and treasure like the long talks, Halloween parties, and breakfasts. If you ever need a scientific question answered or a recommendation for a good book or restaurant, Amena is your person. The breadth of her knowledge in all fields is quite amazing. As a fellow MD/PhD student, Sam has been someone I commiserated with but he has also been a valuable resource when it came to biochemistry, politics, or parenting. Ildar and Cansu, AKA Candar, have been my office buddies for the past few months. They have been an incredible help for discussing my problems, hypotheses, and models. Ildar is a patient teacher and willing to instruct anyone in bioinformatics, small RNA biology, or film theory. Cansu has the ability to brighten anyone's day both literally with her love of pink and figuratively with her bubbly personality. This thesis also would not be possible without the work of Eugenio, who is responsible for most of the bioinformatics analyses. It has been great to get a more computational perspective on my project and I had a lot of fun discussing my project with him. I am also thankful that Eugenio is very good and patient at explaining complex bioinformatics techniques into simple terms that I could understand.

Finally, I have to give a special thank you to my wife, Jess. It's true, that opposites attract and I am scared to imagine where I might be without her organizational skills, patience, and kindness. I am deeply appreciative for her unwavering support and her willingness to sacrifice so many things to help me. While Worcester has treated me very well, I will forever be indebted to my wife for moving

here to be with me rather than staying in California or returning to her home state, Hawaii. Jess truly holds our family together, and both Kai and I are truly lucky to have her in our lives. Before I end, I also wanted to say a quick thank you to Kai, for helping me put everything into perspective and giving me hope for the future.

–Tim

April, 2018

ABSTRACT

Transposons constitute much of the animal genome. While many transposons are ancient and inactivated, numerous others are intact and must be actively repressed. Uncontrolled transposons can cause genomic instability through DNA damage or mutations and must be carefully silenced in the germline or risk sterility or mutations that are passed on to offspring.

In *Drosophila melanogaster*, 23–30 nt long piRNAs direct transposon silencing by serving as guides for Aubergine, Argonaute3, and Piwi, the three fly PIWI proteins. piRNAs derive from piRNA clusters—large heterochromatic DNA loci comprising transposons and transposon fragments. piRNAs are loaded into PIWI proteins via the ping-pong cycle which serves to amplify guide piRNAs. Loaded Piwi then enters the nucleus to transcriptionally repress transposons by establishing heterochromatin. Therefore, to silence transposons, transposon sequences must also be expressed. To bypass this paradox, the HP1 homolog Rhino (Rhi) allows non-canonical, promoter-independent, transcription of transposons embedded in heterochromatin. Transposon RNAs produced in this manner are “incoherent” and have little risk of being translated into transposon-encoded proteins required for transposition.

This thesis focuses on understanding how piRNA clusters permit non-canonical transcription yet restrict canonical transcription. We found that although Rhi promotes non-canonical transcription in piRNA clusters, it also creates a transcriptionally permissive environment that is amenable to canonical transcription. In addition, we discovered that the conserved protein, Maelstrom, is required to repress promoter-driven transcription of individual, potentially active, transposons within piRNA clusters and allows Rhi to transcribe such transposon sequences into incoherent piRNA precursors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

COPYRIGHT INFORMATION

The following publication appears in whole or in part in the thesis:

Chang, T.H., Mattei, E., Weng, Z., and Zamore, P.D. (2018). The Protein Maelstrom Represses Canonical Polymerase II Transcription in Bi-Directional piRNA-producing Loci in *Drosophila melanogaster*. Manuscript in preparation.

LIST OF ABBREVIATIONS

α-Tub: α-Tubulin

A. thaliana: *Arabidopsis thaliana* (thale cress)

Ago: Argonaute

Armi: Armitage

Arx: Asterix

Aub: Aubergine

C. elegans: *Caenorhabditis elegans* (round worm)

ChIP: chromatin immunoprecipitation

Cuff: Cutoff

D. melanogaster: *Drosophila melanogaster* (fruit fly)

D. virilis: *Drosophila virilis* (fruit fly)

Del: Deadlock

dsRNA: double-strand RNA

endo-siRNA: endogenous siRNAs

flam: *flamenco*

H3K4: histone H3 lysine 4

H3K9: histone H3 lysine 9

H3K9me: methylated histone H3 lysine 9

HMT: histone methyltransferase

HP1: Heterochromatin Protein 1

kbp: kilobase pairs

LINE: long interspersed nuclear element

LTR: long terminal repeat

Mael: Maelstrom

miRNA: microRNA

mRNA: messenger RNA

Moon: Moonshiner

nts: nucleotides

OSC: ovarian somatic cell

Panx: Panoramix

PEV: Position effect variegation

PIC: pre-initiation complex

piRNA: Piwi-interacting RNA

Pol II: RNA polymerase II

poly(A): polyadenylation

RDC: Rhino-Deadlock-Cutoff

RDRC: RNA-dependent RNA polymerase complex

RdRP: RNA-dependent RNA-polymerase

RITS: RNA-induced transcriptional gene silencing

Rhi: Rhino

RNAi: RNA interference

S. pombe: *Schizosaccharomyces pombe* (fission yeast)

S. cerevisiae: *Saccharomyces cerevisiae* (budding yeast)

SINE: short interspersed nuclear element

sRNA: small RNA (includes miRNAs, siRNAs, and piRNAs)

siRNA: small interfering RNA

ssRNA: single-strand RNA

Su(var): Suppressor of variegation

T. ni: *Trichoplusia ni* (cabbage looper)

TBP: TATA box-binding protein

TFIIA: basal transcription factor IIA

TIR: terminal inverted repeat

TREX: transcription/export

TRF1: TATA box-binding protein-related factor 1

TSS: transcription start site

TTS: transcription termination site

UTR: untranslated region

w^m: *white mottled*

Zuc: Zucchini

CHAPTER I: INTRODUCTION

HISTORY OF HETEROCHROMATIN

Christening of Heterochromatin

In 1928, Emil Heitz found that chromosomes from different mosses had regions that were more condensed, which he termed “heterochromatin,” and less condensed, which he termed “euchromatin.” Heitz also noted that the amount of heterochromatin varies among chromosomes. In particular, sex chromosomes were consistently enriched with heterochromatin. Moreover, among different cells, certain regions of chromosomes were consistently condensed while other regions tended to vary. These regions would later become known as constitutive and facultative heterochromatin, respectively (Reviewed in (Passarge, 1979)).

Heterochromatin and Gene Expression: Position Effect Variegation

Contemporaneous with Heitz, H. J. Muller discovered that exposing *Drosophila melanogaster* (*D. melanogaster*; fruit fly) to radiation could cause mutations that did not behave like traditional spontaneous lesions (Muller, 1930). Previous work had demonstrated that *D. melanogaster* with an inherited mutation in the *white* gene have a white-eye phenotype, which contrasts their red-eyed wild-type counterparts. Interestingly, after radiation exposure, Muller’s five mutant fly strains did not have a consistent white-eye phenotype, but rather had variegated expression of the *white* gene in the fly compound eye, producing a white-mottled phenotype (w^{m1}–w^{m5}). The lesion that caused the mottled *white* expression was eventually traced to an inversion of the X chromosome at a point near the *white* gene. Moreover, w^{m1} flies displayed an association between the degree of *white* repression and *Notch* expression—flies with eyes that were more mottled also had notched wings, while flies with little mottling had normal wings. Because *Notch* and *white* are physically close to one another on the chromosome (~340 kbp), Muller concluded that gene expression is dependent on its

position on the chromosome, or “chromatin context” (Muller, 1930). This phenomenon was therefore referred to as position effect variegation (PEV; Figure 1.1).

During the next few years, different factors that modified the *white* variegation pattern were discovered, bringing additional insight into PEV. Heterochromatin was identified as one such factor. Researchers observed that the addition of an extra heterochromatic Y chromosome suppressed variegation, while the removal of the Y chromosome enhanced variegation (Gowen and Gay, 1934). In 1936, lesions that caused variegating phenotypes were mapped to heterochromatin (Schultz, 1936), linking the variable inactivation of genes to heterochromatin. However, it was not until 1967, that the first Suppressor of variegation gene (*Su(var)*) was precisely mapped—showing that a single euchromatic gene could affect PEV in *trans* (Spofford, 1967).

Figure 1.1

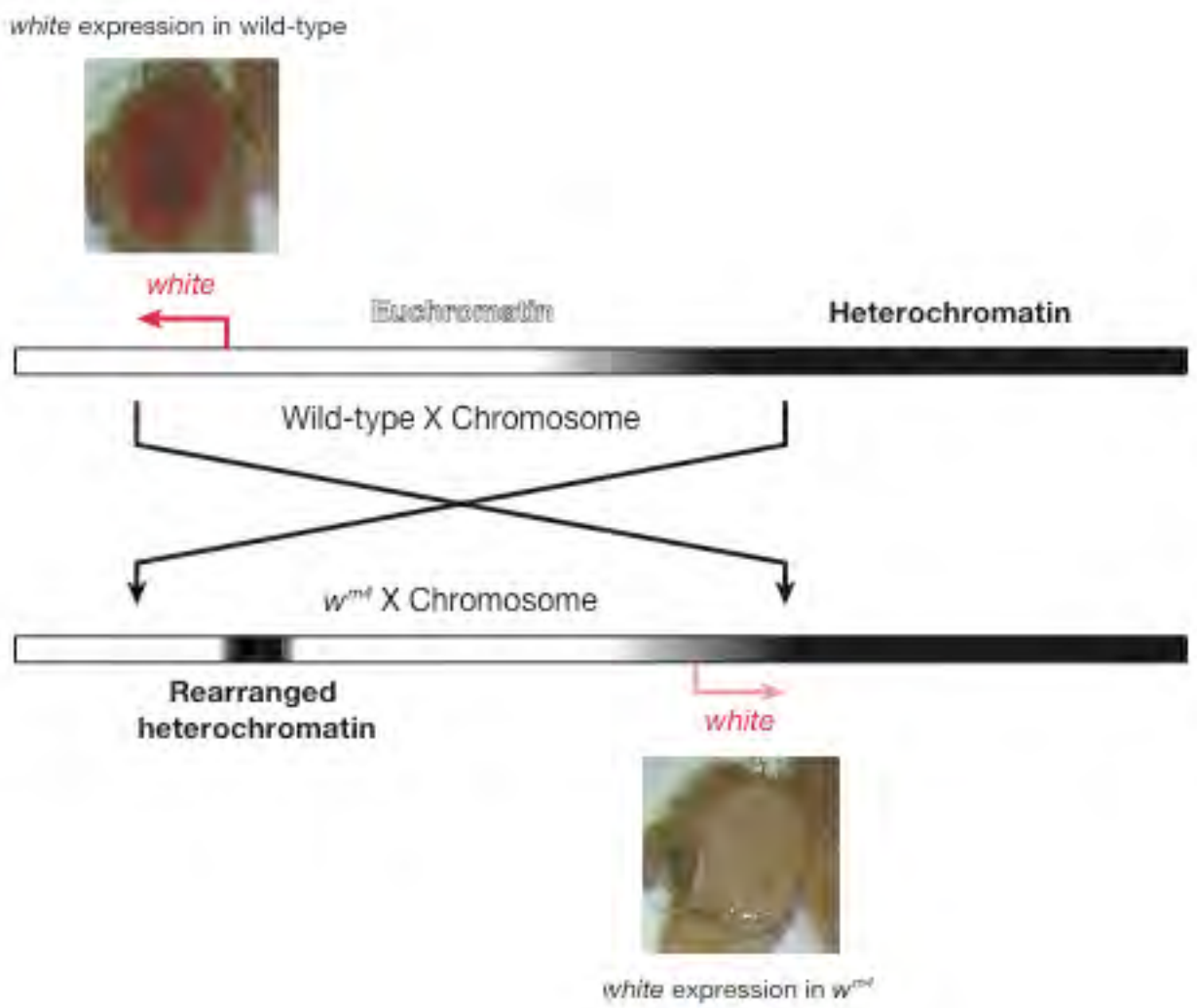

white expression in wild-type
white
Euchromatin
Heterochromatin
Wild-type X Chromosome
w^{mf} X Chromosome
Rearranged heterochromatin
white
white expression in w^{mf}

Figure 1.1: Positon effect rearrangement for *w^{m4}*

In *w^{m4}*, the *white* gene is repositioned next to heterochromatin on the X chromosome which causes the variegated expression. Schematic of the wild-type (top) and *w^{m4}* X chromosome is displayed with representative images of wild-type and *w^{m4}* eyes. Breakpoints of the chromosomal inversion are shown with black arrows and transcription of *white* is depicted as a red arrow. Euchromatin and heterochromatin are shown on the X chromosome as white and black, respectively. Adapted from (Schotta et al., 2003).

Molecular Properties of Heterochromatin

Many characteristics we now associate with heterochromatin formation and silencing were first discovered using PEV. The mechanistic details behind heterochromatin remained a mystery, however, until advances in molecular biology were made. Today, we know that both heterochromatin and euchromatin are composed of the same fundamental unit, the nucleosome, which is made up of an octamer of four conserved histones wrapped around 147 bp of DNA (Luger et al., 1997). The N-terminal tails of histones are thought to be flexible and largely unstructured, allowing for post-translational modifications, such as methylation and acetylation (Luger et al., 1997; Zheng and Hayes, 2003). Post-translational modifications to histone tails are essential for many biological processes ranging from transcription to DNA repair (Harr et al., 2016). Different covalent modifications can change the chemical properties of the histone tail and can act as binding sites for other proteins. Together, multiple modified histone tails may constitute a “histone code” that expands upon the genetic code (Strahl and Allis, 2000; Jenuwein and Allis, 2001; Rando, 2012). Histone methyltransferases (HMTs) are responsible for three common histone tail modifications found in heterochromatin: mono- (H3K9me1), di- (H3K9me2), and tri-methylated (H3K9me3) lysine 9 of histone H3 (H3K9) (Rea et al., 2000; Barski et al., 2007). Thus, modifications making up less than 0.5% of the molecular weight of a histone could potentially determine whether a gene is on or off.

HP1a is Essential to Heterochromatin Function

Perhaps the best-known non-histone chromosomal protein is Heterochromatin Protein 1a (HP1a), which binds to methylated H3K9 (Bannister et al., 2001; Lachner et al., 2001; Nakayama et al., 2001). In flies, HP1a is encoded by *Su(var)2-5*, which as the name suggests, is a suppressor of PEV (Eissenberg et al., 1990). HP1a has three

domains (Figure 1.2; Vermaak and Malik, 2009). The conserved N-terminal chromodomain binds H3K9me2 and gives HP1a its specificity (Paro and Hogness, 1991; Bannister et al., 2001; Lachner et al., 2001; Nakayama et al., 2001; Jacobs and Khorasanizadeh, 2002). The C-terminal chromo shadow domain is also conserved and mediates homodimerization and interactions with other proteins (Aasland and Stewart, 1995; Smothers and Henikoff, 2000). One protein that interacts with the HP1a shadow domain is Su(var)3-9, the HMT responsible for making H3K9me2 (Schotta et al., 2002). The final domain for HP1a is the hinge domain and links the chromo and chromo shadow domains. The hinge domain is the least conserved and is thought to help with HP1a localization by binding to DNA and RNA (Smothers and Henikoff, 2001; Jinek and Doudna, 2008; Meehan et al., 2003).

The chromo shadow domain allows HP1a to dimerize and tether two nucleosomes with H3K9me via the chromodomain. In this manner, HP1a can potentially crosslink many methylated nucleosomes to form a condensed, ordered array that is less accessible to transcription than euchromatin. Moreover, the chromo shadow domain interacts with Su(var)3-9 to methylate nearby H3K9, further spreading the H3K9me2 mark (Dorer and Henikoff, 1994; Fanti et al., 1998; Eissenberg and Elgin, 2000; Hall et al., 2002; Grewal and Elgin, 2002; Cheutin et al., 2003). Thus, HP1a can be thought of like a scaffold protein, acting like a base for which other proteins can interact and localize to heterochromatin (Figure 1.2). In this model, not only does HP1 propagate heterochromatin but is also responsible for repressing gene expression.

Figure 1.2

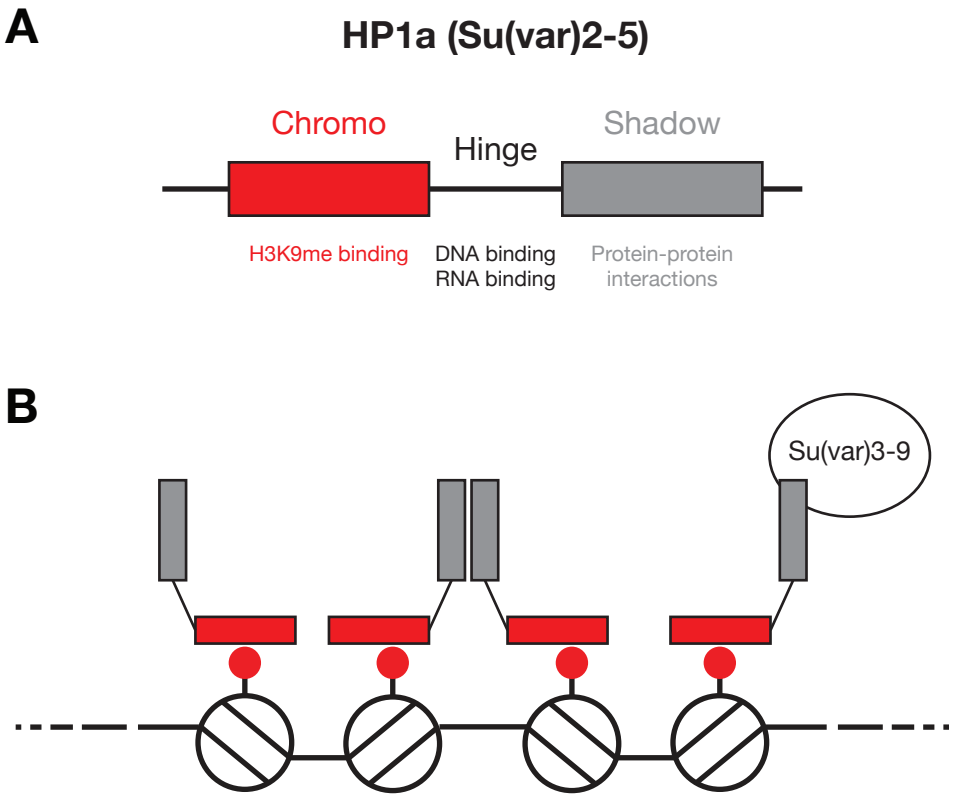

A
HP1a (Su(var)2-5)
Chromo
Hinge
Shadow
H3K9me binding
DNA binding
RNA binding
Protein-protein
interactions
B
Su(var)3-9

Figure 1.2: HP1a maintains heterochromatin

(A) A schematic of HP1a showing its three domains and their activities. (B) A model of how HP1a maintains and propagates heterochromatin through its chromo and chromo shadow domains. Su(var)3-9 is depicted, but other proteins have been identified that interact with the chromo shadow domain. Nucleosomes are depicted as wrapped white circles and methylated H3K9 is depicted as a red circle. Adapted from (Vermaak and Malik, 2009).

Transposons Nucleate Heterochromatin

In 1950, Barbara McClintock noticed that the mutable genes she was studying in maize behaved similarly to PEV in flies and suggested that the two phenomena were related (McCLINTOCK, 1950; McCLINTOCK, 1951). She observed that mutable loci, or "controlling elements" could translocate and give a variegated phenotype and concluded that heterochromatin was likely the controlling factor in this process (McCLINTOCK, 1950). These mutable loci would later be identified as transposons.

In 1984, the connection between PEV, heterochromatin, and transposons came full circle when researchers discovered that the breakpoints for three variegating *white* mutants were transposons. They also observed that when *white* and the surrounding genomic region were reinverted back into euchromatin, the transposon sequences remained heterochromatic. This finding suggests that heterochromatin was actively formed at these transposon sequences and not just propagated from the centromere (Tartof et al., 1984). Further evidence that transposons could act as nucleation sites for heterochromatin was shown when transgenes containing *white* converted from a variegated white-eye phenotype to a red-eye phenotype when a nearby transposon was removed (Sun et al., 2004). Collectively, these early studies suggested that transposons were not only capable of affecting gene expression by altering chromosome structure but could also affect the local expression of genes by spreading heterochromatin.

Heterochromatin Prevents Transposon-Induced Genome Instability

Although McClintock's discovery of mutable genes occurred in 1950, transposon-induced changes in gene expression had already been observed in *Drosophila virilis* (*D. virilis*; fruit fly) as early as 1926 (Demerec, 1926a; Demerec, 1926b). While these early observations gave obvious phenotypes, they did not appear to greatly harm the organism and were thought of as nothing more than genetic oddities. It was only

decades later that transposons began to be considered “selfish” and parasitic (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). For example, flies that cannot repress the *P-element* transposon are sterile (Bingham et al., 1982; Rubin et al., 1982).

Transposons can be classified as Type I or Type II (Slotkin and Martienssen, 2007). Type I elements, or retrotransposons, require a reverse transcription step to transpose, making them similar to retroviruses. Type I elements can be further classified as: long terminal repeats (LTRs), long-interspersed nuclear elements (LINEs), and short-interspersed nuclear elements (SINEs). Type II elements, or DNA elements, do not require a reverse transcription step, but do depend on terminal inverted repeats (TIRs) for transposing. Finally, all transposons can be classified as autonomous and non-autonomous. Autonomous transposons can transpose on their own, but non-autonomous transposons lack essential enzymes and rely on other transposons for mobilization.

The interactions between organisms and transposons are clearly evident in the genome. In eukaryotes, transposons account for ~1% (in the filamentous fungus) to ~85% (in maize) of the genome (Cuomo et al., 2007; Tenaillon et al., 2011). In *D. melanogaster* about 20% of the genome is transposons (Adams et al., 2000; Kaminker et al., 2002). How do organisms cope with such a large number of transposons? One danger with having repetitive sequences in the genome is misalignment during homologous recombination. To prevent this, repetitive sequences are enriched with heterochromatin, which is thought to be recombinationally silent (Peng and Karpen, 2008). One line of evidence that supports this theory is that from plants to insects to mammals, transposons are enriched at pericentromeric heterochromatin (Adams et al., 2000; Arabidopsis, 2000; Gendrel et al., 2002; Kaminker et al., 2002; Martens et al., 2005; Schueler and Sullivan, 2006). Furthermore, transposons enriched with heterochromatin are also transcriptionally repressed. Thus, heterochromatin is essential for maintaining genome stability.

HISTORY OF SMALL RNAs

Discovery of Small RNAs

Like PEV and transposons, the sRNA field also began with a genetic oddity. In 1990, two research groups attempted to increase pigment expression in petunias by adding additional copies of chalcone synthase (CHS), a gene required for pigment production. Rather than the expected darker flowers, they observed the opposite effect: not only did some petals lose pigmentation, but CHS mRNA levels also decreased (Napoli et al., 1990; van der Krol et al., 1990). The phenomenon of silencing a gene when adding additional exogenous copies, or co-suppression, was also found in fungus (Romano and Macino, 1992) and flies (Pal-Bhadra et al., 1997). The mechanism behind co-suppression remained elusive until Andrew Fire and Craig Mello discovered that gene silencing was triggered by double-stranded (dsRNA) (Schrider et al., 2011). Additional insight into post-transcriptional gene silencing came soon after, when the Baulcombe lab determined that small interfering RNAs (siRNAs) could silence genes (Hamilton and Baulcombe, 1999).

siRNAs are loaded into Argonaute (AGO) proteins, which are the effectors of siRNA-mediated silencing or RNA interference (RNAi) (Hutvagner and Simard, 2008; Cenik and Zamore, 2011). The PAZ domain of AGO, a single-stranded-RNA- (ssRNA)-binding module, binds siRNA (typically ~21 nts long), while the PIWI domain possesses the ribonuclease activity required for RNAi (Cerutti et al., 2000). Together, the loaded siRNA specifically guides AGO to its targets through Watson-Crick base pairing for regulation (Figure 1.3; Wee et al., 2012). Both the PAZ and PIWI domains will be covered in greater detail at a later point.

Figure 1.3

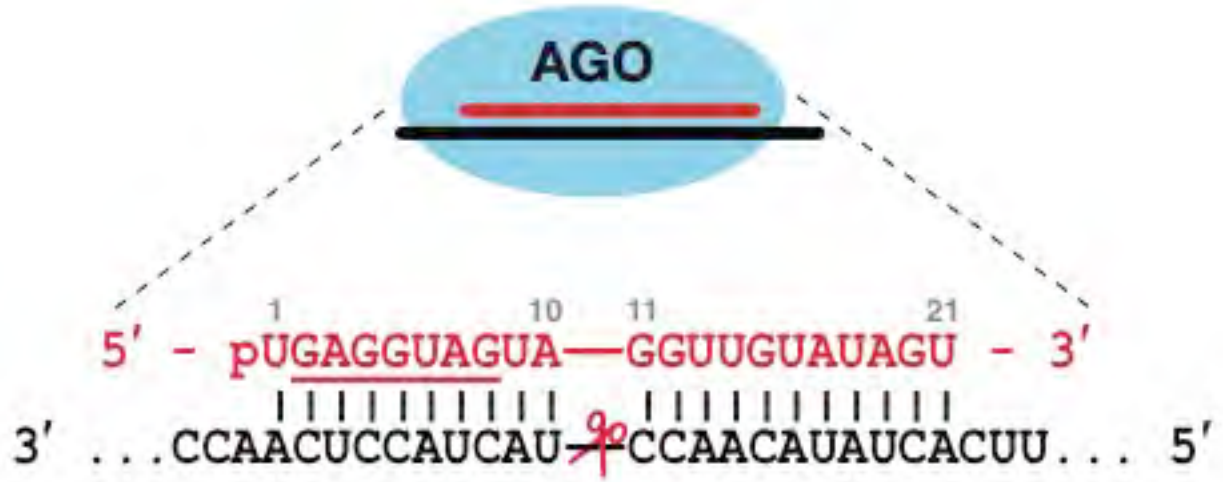

AGO
1
10
11
21
5′ - pUGAGGUAGUA—GGUUGUAUAGU - 3′
3′ . . . CCAACUCCAUCAU CCAACAUAUCACUU . . . 5′

Figure 1.3: Argonaute proteins use sRNAs to find their targets

Schematic of an AGO protein loaded with a guide (red) bound to a target (black) through base pair complementarity. Given sufficient complementarity, AGO proteins can cleave their targets between the 10^{th} and 11^{th} nt of the guide strand (depicted as red scissors).

siRNAs Repress Transposons

The previous examples of co-suppression in flowers, fungi, and flies all demonstrated that exogenous sources of dsRNA, whether injected or produced from a transgene, could direct RNAi. However, hints that RNAi may also have a physiological role came before its discovery when scientists noticed that expression of viral RNA from a transgene appeared to inhibit viral replication (Covey et al., 1997). A few years after the discovery of siRNAs, several labs found that siRNAs in *Arabidopsis thaliana* (*A. thaliana*; thale cress) mapped to endogenous transposons (Hamilton et al., 2002; Llave et al., 2002; Mette et al., 2002; Xie et al., 2004). In *Caenorhabditis elegans* (*C. elegans*; round worm), endogenous transposon-mapping dsRNAs were also detected and mutants resistant to RNAi overexpressed transposons and exhibited transposon mobilization (Ketting et al., 1999; Tabara et al., 1999; Sijen and Plasterk, 2003).

The first reported mammalian endo-siRNAs were found in cultured human cells and mapped to L1 retrotransposons (Yang and Kazazian, 2006). With the advent of deep-sequencing, small RNAs from mouse oocytes were later identified and endo-siRNAs mapping to transposons were also detected. Repression of these transposons was also dependent on siRNAs (Tam et al., 2008; Watanabe et al., 2008). Deep sequencing also allowed detection of transposon-mapping endo-siRNAs in germline and somatic tissues of *D. melanogaster* (Czech et al., 2008; Chung et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008). Therefore, one conserved function for RNAi is silencing endogenous transposons.

siRNAs Produced from Heterochromatin Assemble Heterochromatin

S. pombe centromeres, like other eukaryotic centromeres, are heterochromatic and repetitive (Clarke and Baum, 1990). When a transgene was inserted within the centromere, individual colonies had a distinct, variegated pattern resembling PEV

(Allshire et al., 1994). While *S. pombe* does not appear to have siRNAs that map to transposons, RNAi mutants have modest transposon overexpression (Hansen et al., 2005). Most siRNAs correspond to centromeric repeats, and loss of RNAi led to the loss of repression of centromeric transcripts and transgenes (Provost et al., 2002; Reinhart and Bartel, 2002; Volpe et al., 2002; Cam et al., 2005; Djupedal et al., 2009). Analysis of nascent transcripts revealed that centromeric repeats were normally transcribed from both strands but only detected in RNAi mutants, suggesting they were rapidly processed into siRNAs and loaded into Ago1 (Volpe et al., 2002). Interestingly, loss of Ago1 or siRNAs led to a large reduction in H3K9 methylation at centromeric repeats (Volpe et al., 2002). Ago1 was later discovered to be part of the RNA-induced transcriptional silencing (RITS) complex, which is required for heterochromatin assembly (Verdel et al., 2004). Therefore, transcription of centromeric repeats by RNA polymerase II (Pol II) is required for heterochromatin assembly at centromeric repeats (Kato et al., 2005; Djupedal et al., 2005). This seemingly paradoxical concept begs the question: *how does an organism silence a sequence when the transcription of the sequence is required for silencing?*

To bypass this conundrum, *S. pombe* transcribes centromeric repeats during early S-phase, when heterochromatin is more accessible (Smith et al., 1995; Kim et al., 2003; Chen et al., 2008; Kloc et al., 2008). During transcription, double-stranded siRNA precursors (Bernstein et al., 2001; Knight and Bass, 2001; Ketting et al., 2001) are produced by the RNA-dependent RNA polymerase complex (RDRC) (Motamedi et al., 2004). Interestingly, the RDRC and RITS complexes physically interact (Motamedi et al., 2004) and can be targeted to nascent transcripts by siRNAs in the RITS complex (Verdel et al., 2004; Petrie et al., 2005; Bühler et al., 2006). Thus, both RITS/RDRC complexes localize at nascent transcripts to produce more siRNAs that can further promote heterochromatin formation in *cis*. This model suggests a self-reinforcing positive amplification loop that explains the maintenance and potential spread of

centromeric heterochromatin in *S. pombe* (Noma et al., 2004; Sugiyama et al., 2005; Irvine et al., 2006).

Like *S. pombe*, *A. thaliana* employs a similar siRNA-dependent feed-forward loop to promote heterochromatin formation. Unlike *S. pombe*, *A. thaliana* bypass heterochromatin silencing using two plant-specific nuclear RNA polymerases, Pol IV and Pol V. Pol IV localizes to heterochromatin because of its interaction with SHH1, which binds methylated H3K9 (Kowalczykowski et al., 1980; Zhang et al., 2013), and initiates siRNA biogenesis in association with an RdRP (Herr et al., 2005; Onodera et al., 2005; Blevins et al., 2015; Zhai et al., 2015). Pol V transcripts act as scaffolds and are targeted by siRNAs loaded into AGO4 (Wierzbicki et al., 2009) which can form a complex with a *de novo* methyltransferase (Zhong et al., 2014). Methylated DNA is then recognized by a chromatin-remodeling complex that includes two Su(var)3-9 H3K9 HMT homologs (Liu et al., 2014). Completing the cycle, methylated H3K9 allows Pol IV to transcribe siRNAs that are complementary to Pol V transcripts. In summary, both fission yeast and plants have evolved mechanisms to safely transcribe potentially dangerous transposons by coupling transcription to siRNA production, thus destroying the transcript. Furthermore, in both species, siRNA creation is coupled to heterochromatin formation to ensure that transcription of harmful sequences is repressed except under carefully regulated circumstances.

A similar phenomenon appears to occur in flies, where RNAi components were found to be required for transcriptional silencing and were also suppressors of PEV (Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004). Loss of these components caused the loss of H3K9me and delocalization of HP1a (Pal-Bhadra et al., 2004). Therefore, like *S. pombe* and *A. thaliana*, RNAi also appeared to assemble heterochromatin in *D. melanogaster*. Two genes involved in heterochromatin formation, *piwi* and *aubergine*, would soon become central to a new class of sRNAs: piRNAs.

THE piRNA PATHWAY

Discovery of piRNA Genes

D. melanogaster has long been used as a model organism to study developmental embryo patterning. Mutations that affect either the anterior-posterior or dorsal-ventral axes can be easily screened for patterning defects using a microscope (Schupbach and Wieschaus, 1986). Using this method, many “maternal effect” genes that were required for axis specification were discovered (Schüpbach and Wieschaus, 1989; Schüpbach and Wieschaus, 1991). Maternal effect genes often encode mRNA or protein that the mother deposits in the oocyte. Because the early embryo is transcriptionally inactive in many organisms, the maternal supply is especially important and mutations to maternal genes affect the offspring (Figure 1.4; Schier, 2007). Many of the candidate genes found in these initial studies would later be found to be in the piRNA pathway, including: *aubergine*, *cutoff*, *deadlock*, *vasa*, and *zucchini* (Schüpbach and Wieschaus, 1989; Schüpbach and Wieschaus, 1991).

Figure 1.4

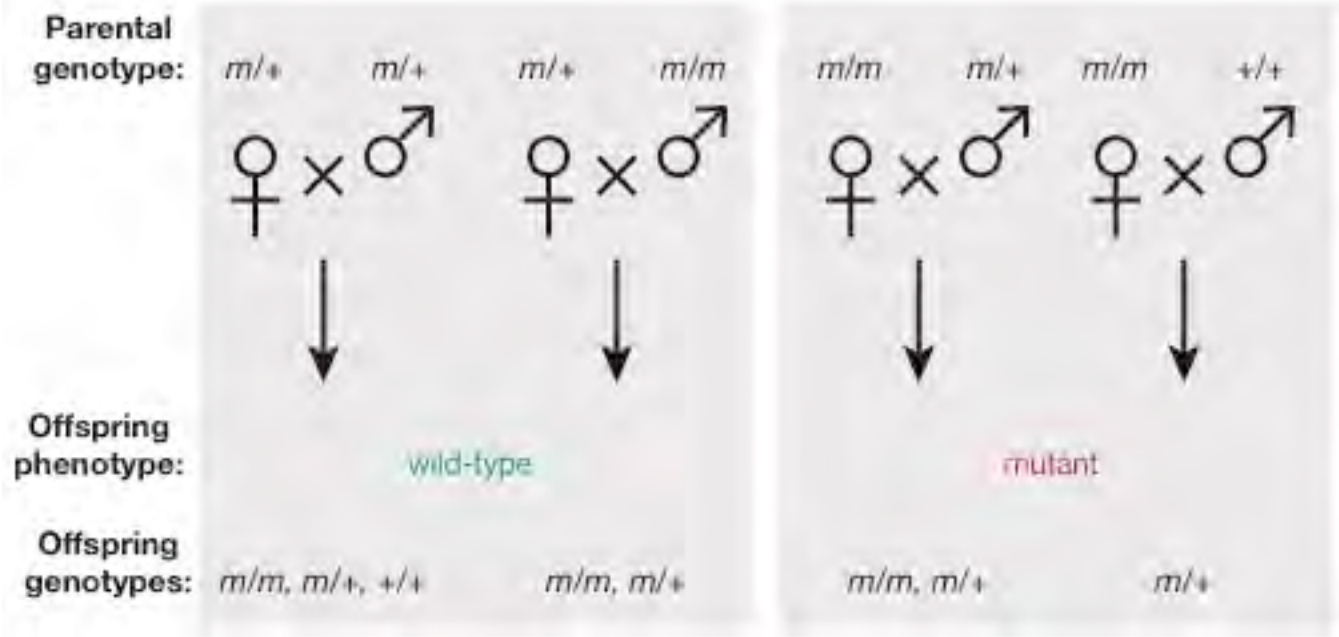

Parental genotype:
m/+
m/+
m/+
m/m
m/m
m/+
m/m
+/+
Offspring phenotype:
wild-type
mutant
Offspring genotypes:
m/m, m/+, +/+
m/m, m/+
m/m, m/+
m/+

Figure 1.4: Maternal effect phenotypes determined by mother

Genetic crosses involving a maternal effect recessive mutation (*m*). For maternal effect genes, the phenotype of the offspring is determined by the genotype of its mother. Thus, offspring of the same genotype may have different phenotypes. For many piRNA pathway genes, maternal effect phenotypes can be observed during oogenesis.

piRNAs, A New Class of sRNAs

In 2001, a class of longer siRNAs (25–30 nt), called repeat-associated small interfering RNAs (rasiRNAs), was discovered in *D. melanogaster* testes (Aravin et al., 2001). Similar to siRNAs from other organisms, rasiRNAs were found to map to transposons and repetitive regions, and their loss led to elevated transposon expression (Aravin et al., 2001; Aravin et al., 2003; Aravin et al., 2004; Sarot et al., 2004). After discovering that rasiRNA biogenesis was distinct from siRNAs, researchers realized these RNAs were not actually siRNAs at all, and thus, were renamed Piwi-interacting RNAs (piRNAs) (Girard et al., 2006). Rather than dsRNA precursors, piRNAs derive from ssRNA precursors and are Dicer-independent (Pelisson et al., 2007; Saito et al., 2006; Vagin et al., 2006). A conserved feature of piRNAs is that they are 2′-*O*-methylated at their 3′ termini by Hen1. In flies, siRNAs are also 2′-*O*-methylated but not miRNAs (Vagin et al., 2006; Horwich et al., 2007; Lagarrigue et al., 2013; Pelisson et al., 2007; Saito et al., 2007).

piRNAs are found in diverse animals from the simple *Amphimedon queenslandica* (sea sponge) to humans (Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006; Grimson et al., 2008). Thus, piRNAs emerged before the divergence of bilaterian animals. Like flies, sea sponge and mammals all have piRNAs that map to transposons, although in mammals piRNAs that map to genes and intergenic regions were much more abundant (Grivna et al., 2006a; Aravin et al., 2007; Girard et al., 2006; Grimson et al., 2008). Finally, as opposed to miRNAs and siRNAs, piRNA sequences are extremely diverse. While piRNAs often map to multiple genomic loci, many piRNA species are only sequenced once (Brennecke et al., 2007).

PIWI Proteins

AGO and PIWI proteins are the two sub-families of the Argonaute family proteins. The two sub-families share three key conserved domains: PIWI, MID, and PAZ (Cenik and Zamore, 2011). The PIWI domain resembles the RNase H nuclease and can cleave the phosphodiester bond of the target RNA between the 10^{th} and 11^{th} nt of guide RNA (Liu et al., 2004; Song et al., 2004). The MID domain anchors the 5′ monophosphate of the bound sRNA by forming a binding pocket and helps the sRNA pair with target RNA (Ma et al., 2005; Wang et al., 2008; Boland et al., 2011). Finally, the PAZ domain anchors the 3′ end of the bound sRNA and is different between AGO and PIWI proteins depending on the 3′ modification of the sRNA (2′ hydroxl for miRNAs and 2′-*O*-methyl for piRNAs and fly siRNAs) (Ma et al., 2004; Lingel et al., 2004; Kawaoka et al., 2011; Tian et al., 2011).

As the differences in the PAZ domain suggest, AGO proteins are loaded with miRNAs and siRNAs while PIWI proteins are loaded with piRNAs. AGO and PIWI proteins are further distinguished from each other by their localization. AGO proteins are found in all cells while PIWI proteins tend to be enriched in the gonads of many animals, including *D. melanogaster*. Recently, somatic piRNAs have also been found in many arthropod species including *D. virilis*, which suggests that *D. melanogaster* and closely related species of the *melanogaster* group may actually be the exception to the rule among insects (Lewis et al., 2018).

D. melanogaster have three PIWI proteins Piwi, Aubergine (Aub), and Argonaute3 (Ago3). All three PIWI proteins are non-redundant and are required for fertility and germline transposon repression (Cox et al., 1998; Wilson et al., 1996; Vagin et al., 2006; Saito et al., 2006; Brennecke et al., 2007; Gunawardane et al., 2007; Li et al., 2009a; Malone et al., 2009). In fact, the axis specification defects observed in mutations of piRNA pathway genes are caused by the loss of transposon repression.

Active transposons can cause double-strand breaks that activate the DNA damage signaling pathway and lead to patterning defects (Klattenhoff et al., 2007).

Aub and Ago3 are found in germ cell cytoplasm and enriched at the perinuclear structure called the nuage (French for “cloud; Eddy, 1974. In flies, using an electron microscope, the nuage appears as nebulous, electron-dense bodies surrounding the nucleus. Many other proteins required for piRNA production also localize to the nuage including Vasa, Armitage (Armi), Zucchini (Zuc), Krimper, and Qin (Hay et al., 1988; Cook et al., 2004; Lim and Kai, 2007; Pane et al., 2007; Zhang et al., 2011). The localization of these proteins to nuage all depend on Vasa, as *vasa* mutants appear to lose nuage and therefore all the other components of nuage are also lost (Liang et al., 1994; Lim and Kai, 2007).

In contrast, Piwi, the only nuclear PIWI protein, acts in both the germline and adjacent somatic follicle cells and represses transposon transcription rather than cleaving their transcripts (Cox et al., 2000; Brennecke et al., 2007; Malone et al., 2009; Klenov et al., 2011). Interestingly, while the catalytic function of Piwi is conserved, it is not required for transposon repression (Saito et al., 2009; Sienski et al., 2012; Ronsseray et al., 1984). In somatic follicle cells, Piwi is loaded with piRNAs produced from *flamenco* (*flam*), a 180 kbp heterochromatic pericentromeric locus, and silences *gypsy* retrotransposons, which can infect and integrate into neighboring germ cell DNA to be passed on to the next generation (Pelisson et al., 1994; Prud'homme et al., 1995; Song et al., 1997; Chalvet et al., 1999; Robert et al., 2001; Sarot et al., 2004; Mevel-Ninio et al., 2007; Pelisson et al., 2007)

Identification of *D. melanogaster* piRNA Clusters

Analysis of *D. melanogaster* transposon insertions predicted the existence of a “co-suppression network that may act as a global surveillance system” against transposons

(Bergman et al., 2006). Furthermore, deep-sequencing technology allowed identification of many new species of piRNAs that mapped to distinct genomic loci called “piRNA clusters” (Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006; Brennecke et al., 2007). In *D. melanogaster*, ~140 piRNA clusters produce 81% of all piRNAs that uniquely map to the genome yet comprise only 3.5% of the genome (Brennecke et al., 2007). piRNA clusters are also enriched near pericentromeric and telomeric heterochromatin and are made up of transposon sequences that are inactivated by mutations or because they are fragmented by insertions of other transposons (Figure 1.5).

Figure 1.5

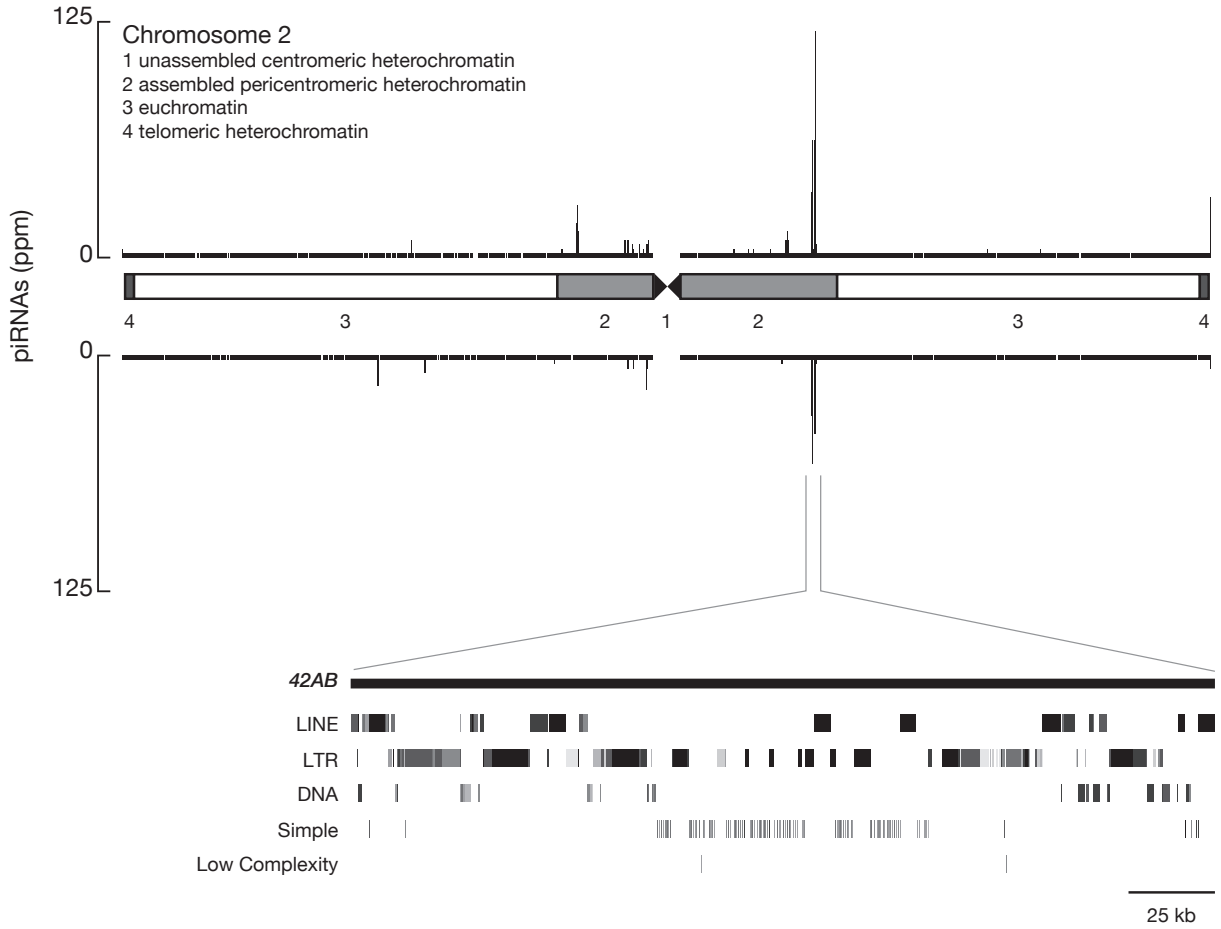

125
Chromosome 2
1 unassembled centromeric heterochromatin
2 assembled pericentromeric heterochromatin
3 euchromatin
4 telomeric heterochromatin
piRNAs (ppm)
0
4
3
2
1
2
3
4
0
125
42AB
LINE
LTR
DNA
Simple
Low Complexity
25 kb

Figure 1.5: piRNAs map to discrete loci

D. melanogaster piRNAs map to discrete loci called “piRNA clusters” which comprise transposons and transposon fragments. Schematic of wild-type piRNAs overlaid onto chromosome 2. Major chromosome domains are labeled 1–4. RepeatMasker annotations for *42AB* are shown in higher resolution below. Adapted from (Brennecke et al., 2007).

Transcription of piRNA Clusters

In flies, some piRNA clusters are transcribed from one genomic strand and are called uni-strand clusters (Brennecke et al., 2007). These clusters include *cluster2* and the somatic cluster, *flam*. Although *flam* is large, heterochromatic, and composed of transposons, it is transcribed by conventional, promoter-initiated, Pol II transcription that generates spliced, polyadenylated, precursor piRNAs (Robert et al., 2001; Mevel-Ninio et al., 2007; Goriaux et al., 2014). Furthermore, *flam* is unusual among clusters in that transposon fragments are arranged in the genome predominantly in one orientation and transcribed in the antisense orientation. This ensures that *flam* transcripts are non-coding transposon sequences and that the piRNAs produced from *flam* are complementary to active somatic transposons.

On the other hand, dual-strand piRNA clusters are transcribed from both genomic strands and are the source for most of the piRNAs in the germline. The largest dual-strand cluster, *42AB*, produces about 30% of all germline piRNAs (Brennecke et al., 2007). Unlike uni-strand clusters, dual-strand clusters generally lack conserved promoters and chromatin immunoprecipitation (ChIP-) seq did not reveal H3K4me2 or Pol II enrichment in dual-strand clusters—both markers of promoter-initiated transcription (Mohn et al., 2014). Furthermore, dual-strand cluster transcription initiation does not appear to depend on read-through transcription from flanking genes (Chen et al., 2016; Andersen et al., 2017). Instead, Cap-sequencing, which specifically sequences RNAs with 5′ 7-methylguanylate caps, revealed many potential transcription start sites (TSSs) on both genomic strands in dual-strand piRNA clusters (Andersen et al., 2017). This suggests that Pol II initiates within piRNA clusters.

Rather than typical promoter-initiated transcription, fly dual-strand clusters require Rhino (Rhi), an HP1 homolog, to facilitate transcription of dual-strand clusters (Cogoni and Macino, 1999; Le Thomas et al., 2014; Mohn et al., 2014). Consistent with

non-canonical transcription at dual-strand clusters, Rhi suppresses splicing and allows transcription to continue past transcription termination sites (TTSs) (Mohn et al., 2014; Zhang et al., 2014). Furthermore, very few dual-strand transcripts are polyadenylated (Le Thomas et al., 2014). Because *rhi* mutants lose dual-strand transcripts, they cannot make piRNAs to repress transposons and are sterile (Klattenhoff et al., 2009; Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014). Like other HP1 proteins, the Rhi chromodomain preferentially binds H3K9me3 (Le Thomas et al., 2014; Mohn et al., 2014; Yu et al., 2015a). This is consistent with the report that piRNA production requires trimethylation of H3K9 by dSETDB1 (Rangan et al., 2011). dSETDB1 functions in germline-stem cells and as germline cysts differentiate, methylation is gradually taken over by Su(var)3-9 (Yoon et al., 2008). Interestingly, ovaries without dSETDB1 are sterile while Su(var)3-9 is dispensable for fertility (Tschiersch et al., 1994; Yao et al., 2012). This suggests that piRNA cluster formation and transcription is required early in oogenesis.

In addition, like HP1a, Rhi acts like a scaffold, likely mediated by its chromo shadow domain, to tether other heterochromatic dual-strand clusters (Mohn et al., 2014). Two proteins that colocalize with Rhi are Cutoff (Cuff) and Deadlock (Del), which together form the RDC complex (Pane et al., 2011; Le Thomas et al., 2014; Mohn et al., 2014). Like Rhi, both Cuff and Del are also required for dual-strand transcription and piRNAs, and loss of either results in transposon expression and sterility (Wehr et al., 2006; Chen et al., 2007; Pane et al., 2011; Czech et al., 2013; Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014; Chen et al., 2016). Rhi, Cuff, and Del are also interdependent: loss of Rhi, Cuff, or Del leads to the delocalization of the other two proteins (Mohn et al., 2014).

Del does not have any conserved domains but interacts with the Rhi chromodomain and is thought to act as a flexible linker (Mohn et al., 2014). One protein that co-immunoprecipitates (co-IPs) with Del and colocalizes with the RDC is

Moonshiner (Moon; (Andersen et al., 2017). Moon is a paralog of a basal transcription factor IIA (TFIIA) subunit. In eukaryotes, transcription initiation is a conserved stepwise process in which general transcription factors help assemble and stabilize Pol II at promoter DNA (Buratowski et al., 1989; Sainsbury et al., 2015). First, TATA box-binding protein (TBP), a subunit of TFIID, binds to the TATA box, which is located ~20-30 bp upstream of the TSS in flies. While about 40% of *D. melanogaster* promoters have canonical TATA boxes, TBP binding and interaction is not specific to TATA boxes but likely requires additional factors (Kutach and Kadonaga, 2000; Blair et al., 2012; Rhee and Pugh, 2012). Next, TFIIA stabilizes the TBP-DNA complex to form a committed pre-initiation complex (PIC; Imbalzano et al., 1994; Lieberman and Berk, 1994; Papai et al., 2010). TFIIB is also required for Pol II assembly at the promoter and facilitates and stabilizes TBP binding to DNA (Ha et al., 1991; Zhao and Herr, 2002).

Like TFIIA, Moon is required for transcription initiation. Rather than TBP, Moon interacts with a short isoform of TBP-related factor 2 (TRF2), an animal TBP paralog that is expressed during embryogenesis and required for fertility, and TFIIA-S to form an alternative TFIIA complex at dual-strand clusters. (Dantonel et al., 2000; Kaltenbach et al., 2000; Veenstra et al., 2000; Martianov et al., 2001; Zhang et al., 2001; Kopytova et al., 2006; Andersen et al., 2017). Therefore, instead of canonical promoter-dependent transcription initiation, dual-strand clusters require the RDC complex to guide chromatin-dependent transcription. Because Moon does not require a promoter to function, transcription can initiate from either genomic strand, giving dual-strand clusters their defining characteristic.

Although *cluster38C1* is a dual-strand cluster, it does not require Moon for transcription initiation because it also has encoded flanking promoters (Mohn et al., 2014; Chen et al., 2016; Andersen et al., 2017). Furthermore, in *moon* mutants, piRNAs can be detected over 10 kbp downstream of the promoter in the same orientation

(Andersen et al., 2017). This suggests that dual-strand cluster transcripts can be over 10 kb in length.

Cuff, the other member of the RDC complex, is related to the Rai1/DXO/Dom3Z decapping enzyme (Baulcombe, 1999; Chang et al., 2012; Jiao et al., 2013). In *S. cerevisiae*, Rai1 enhances Rat1 5′-3′ exoribonuclease activity, which is important for Pol II termination by degrading the nascent RNA downstream of the poly(A) cleavage site (Xue et al., 2000; Kim et al., 2004). Under the "torpedo model" of transcription termination, Pol II transcription termination involves two steps. First, when transcription reaches the poly(A) sequence, the cleavage and polyadenylation specificity factor (CPSF) complex cleaves the pre-mRNA while Pol II continues transcription. Second, the uncapped residual RNA is then degraded by the 5′-3′ exoribonuclease Rat1/XRN2 until it overtakes knocks off Pol II.

Cuff promotes dual-strand piRNA cluster transcription by suppressing Pol II termination and protecting uncapped dual-strand cluster transcripts from nuclear degradation and evidence suggests that Cuff prevents CPSF from binding and cleaving the nascent dual-strand transcripts at poly(A) sequences (Chen et al., 2016). Therefore, Cuff allows transcription to continue past TTSs and may partially explain how *cluster38C1* produces transcripts over 10 kbp long in *moon* mutant ovaries (Andersen et al., 2017). However, even with Cuff, a significant fraction of cluster transcripts is still cleaved (Chen et al., 2016). Read-through transcripts created by CPSF cleavage have 5′ monophosphates and are targets for Rat1/XRN2 exonuclease activity. Cuff prevents their degradation by stabilizing uncapped transcripts (Chen et al., 2016). Finally, Rhi-dependent splicing suppression is mediated by Cuff (Zhang et al., 2014; Chen et al., 2016). Because uncapped RNAs cannot bind the nuclear cap-binding complex, which promotes splicing (Patzelt et al., 1987; Izaurralde et al., 1994; Lewis et al., 1996), the authors suggest that Cuff may suppress splicing indirectly by stabilizing uncapped transcripts that are less likely to be spliced (Chen et al., 2016). Altogether, Rhi localizes

Cuff to dual-strand clusters where Cuff is able to promote read-through transcription and protect abnormal dual-strand transcripts from degradation. A model for the transcription of dual-strand piRNA clusters is presented in Figure 1.6.

Export of Dual-Strand piRNA Cluster Transcripts

Different steps in pre-mRNA processing are coupled, including the export of the mRNA (Hirose and Manley, 2000; Bentley, 2002; Bentley, 2005). Moreover, many proteins involved with pre-mRNA processing are also conserved between yeast and metazoans, including the TREX (transcription/export) complex which includes UAP56 and the THO complex (Reed and Cheng, 2005). UAP56 is a ubiquitously expressed DEAD box RNA helicase that regulates splicing and mRNA nuclear export (Shen, 2009). In metazoans, loading of the multi-subunit THO complex to nascent transcripts is dependent on pre-mRNA splicing (Masuda et al., 2005). Interestingly, UAP56, Thoc5, and other members of the TREX complex interact with dual-strand cluster transcripts, which are not spliced, and are required for dual-strand piRNA production (Zhang et al., 2012; Hur et al., 2016; Zhang et al., 2014). In germline nurse cells, this is accomplished through Cuff, which interacts with Thoc5 and is required for localization of both Thoc5 and UAP56 to nuclear foci (Hur et al., 2016). In this model, Cuff, which is responsible for suppressing splicing, loads TREX onto nascent dual-strand cluster transcripts (Figure 1.6). Then UAP56, which also colocalizes with Rhi, associates with nascent dual-strand cluster transcripts helps to export the piRNA precursor through the nuclear pore to Vasa, which is in the nuage, to be processed into piRNAs (Zhang et al., 2012; Hur et al., 2016; Zhang et al., 2014).

Figure 1.6

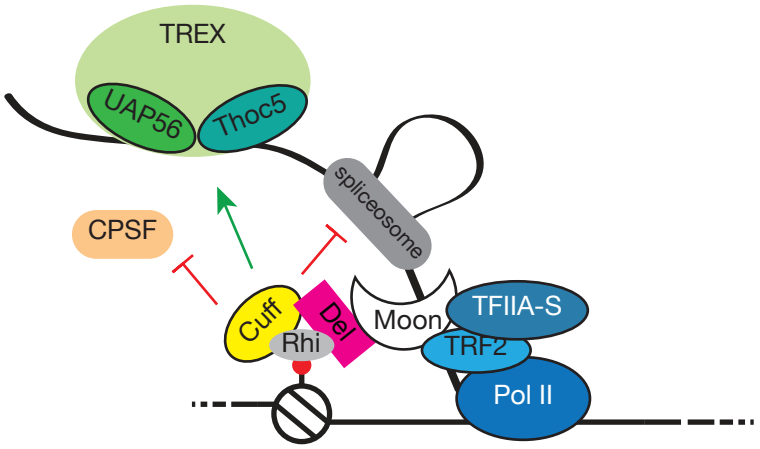

TREX
UAP56
Thoc5
spliceosome
CPSF
Cuff
Del
Rhi
Moon
TFIIA-S
TRF2
Pol II

Figure 1.6: Transcription of dual-strand piRNA clusters

In flies, non-canonical transcription of dual-strand clusters is mediated by Rhi and initiated by Moon independent of promoters. Moon interacts with the Rhi-Del-Cuff (RDC) complex through Del, and is thought to form an alternative TFIIA complex with TRF2 and TFIIA-S. The RDC complex also suppresses splicing, stabilizes uncapped transcripts, and prevents transcription termination. Transcription of dual-strand clusters is also coupled to export and the RDC is required for loading the TREX complex with piRNA precursors. Adapted from (Huang et al., 2017).

piRNA Precursors are Processed into Mature piRNAs in the Cytoplasm

After a piRNA precursor enters the nuage, the 5′ end of the piRNA is produced by slicer-mediated cleavage by Aub or Ago3. Aub-loaded piRNAs correspond mostly to antisense transposon sequences and therefore Aub cleaves sense transposon sequences to generate a 5′ monophosphorylated RNA. This RNA is then loaded into Ago3 to produce a new sense transposon-mapping piRNA that can cleave a complementary antisense transposon sequence to be loaded into Aub. This feed forward amplification loop of RNA slicing and piRNA production is called the “ping-pong cycle.” piRNAs produced by ping-pong can be detected by their characteristic 10 bp overlap between the 5′ ends of sense and antisense piRNA pairs (Brennecke et al., 2007; Gunawardane et al., 2007).

While many of the proteins required for ping-pong are found in the nuage, several proteins that are also required for piRNA processing are found near the outer membrane of the mitochondria. Armi, an RNA helicase, is found in both the nuage and mitochondria while Zuc, an endonuclease, is localized to mitochondria. Both are required for phased piRNA production (Cook et al., 2004; Pane et al., 2007; Malone et al., 2009; Handler et al., 2011; Czech et al., 2013; Han et al., 2015; Mohn et al., 2015; Pandey et al., 2017; Rogers et al., 2017). In the germline, phased piRNAs are initiated by Aub or Ago3 cleavage of a piRNA precursor, generating the 5′ end (Han et al., 2015; Mohn et al., 2015; Wang et al., 2015). Zuc generates the 3′ end of piRNAs produced by ping-pong and can continue downstream in 5′ to 3′ direction generating the 3′ end of the upstream and the 5′ end of the downstream piRNA that is mainly loaded into Piwi (Han et al., 2015; Mohn et al., 2015).

In the ping-pong cycle, sense piRNAs loaded into Ago3 generate antisense piRNAs that can post-transcriptionally silence active transposons. However, for most germline transposons, Piwi plays a larger role in repression than Aub or Ago3 (Wang et

al., 2015). Ago3 is particularly important for loading Piwi because cleavage of antisense transposon sequences initiates the production and defines the identities of phased Piwi-bound piRNAs (Senti et al., 2015; Wang et al., 2015). To ensure that Piwi is enriched with antisense piRNAs, Qin prevents products of Aub cleavage (usually sense transposon sequences) from being loaded into Piwi (Zhang et al., 2011; Wang et al., 2015). Moreover, loss of Aub and Ago3 results in the loss of Piwi-bound piRNAs and the inability of Piwi to enter the nucleus where it represses transposons (Figure 1.7; Wang et al., 2015).

Though less efficient, piRNA 3′ ends can also be created by Aub or Ago3 cleavage and require trimming by the 3′ to 5′ exoribonuclease, Nibbler (Hayashi et al., 2016). The final step of piRNA maturation is methylation of the 2′ hydroxyl group of the 3′ terminus by Hen1 (Horwich et al., 2007; Saito et al., 2007; Wang et al., 2016).

Figure 1.7

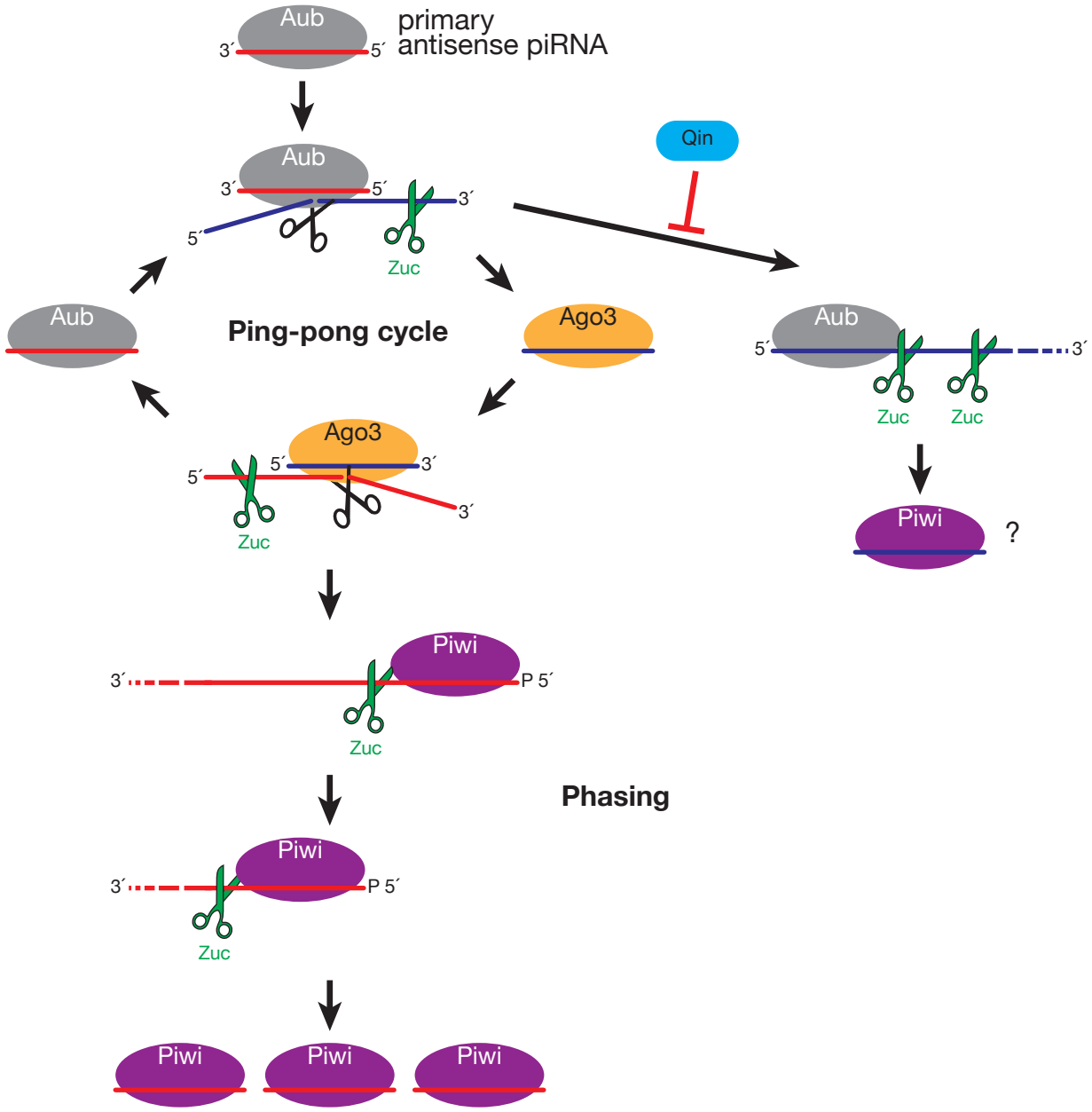

Aub
primary
antisense piRNA
3´
5´
Aub
Qin
Zuc
Ping-pong cycle
Aub
Ago3
Aub
Zuc
Zuc
Piwi
?
Ago3
Zuc
Piwi
P 5´
Zuc
Phasing
Piwi
P 5´
Zuc
Piwi
Piwi
Piwi

Figure 1.7: piRNA biogenesis in the fly ovarian germline

Primary antisense (red) piRNAs that are maternally deposited or produced de novo enter the ping-pong cycle by slicing (black scissors) transposon mRNAs to generate the 5′ end of a secondary sense (blue) piRNAs which are loaded into Ago3. Zuc (green scissors) cleaves downstream and defines the 3′ end of the upstream piRNA and the 5′ end of the downstream piRNA. Qin prevents Aub from getting loaded with sense piRNAs and reinforces heterotypic ping-pong. Ago3, loaded with a sense piRNA, can then cleave an antisense cluster transcript and generate the 5′ end of a new piRNA that can be loaded into Aub or Piwi. Afterwards, Zuc can processively cut the transcript every ~27 nt to generate phased piRNAs. Adapted from (Wang et al., 2015).

Piwi-Mediated Transcriptional Silencing

Rather than cleaving, Piwi is thought to transcriptionally repress transposons by binding to nascent transposon mRNA through its piRNA guide (Aravin et al., 2008; Saito et al., 2009; Shpiz et al., 2011; Sienski et al., 2012; Ronsseray et al., 1984; Le Thomas et al., 2013; Post et al., 2014). Therefore, Piwi must enter the nucleus to function but not before it is loaded with a piRNA (Klenov et al., 2011; Le Thomas et al., 2013). In the nucleus, Piwi, with the help of other proteins, assembles heterochromatin (Sienski et al., 2012; Donertas et al., 2013; Le Thomas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013; Rozhkov et al., 2013; Klenov et al., 2014; Sienski et al., 2015; Yu et al., 2015b). One of these proteins is the zinc finger protein, Asterix (Arx, also known as Gtsf1). While its mechanism is still unknown, Arx is not required for piRNAs but directly interacts with Piwi to establish H3K9me3 at transposons (Donertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013). Another protein that functions downstream of Piwi is Panoramix (Panx, also known as Silencio). Panx associates with Piwi that is bound to transposons and induces heterochromatin formation through its interaction with dSETDB1, which methylates H3K9 (Sienski et al., 2015; Yu et al., 2015b). Because HP1a is integral in heterochromatin formation and maintenance, germline transposons repressed by HP1a are similar those repressed by Piwi (Wang and Elgin, 2011).

Most H3K9me3 peaks do not overlap Rhi peaks (Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014). Interestingly, Piwi is not only able to repress euchromatic transposons through H3K9me3 formation but is also required for localizing Rhi to some of these transposons, turning them into piRNA source loci (Mohn et al., 2014). These two functions are likely closely linked because incoherent Rhi-dependent transcription effectively silences the transposon (Andersen et al., 2017; Zamore, 2017). Furthermore, because dual-strand clusters transcribe both sense and antisense transposon sequences, Piwi-bound piRNAs should be able to bind nascent cluster

transcripts as well. Despite Piwi functioning upstream of dSETDB1, which is required for cluster transcription, loss of nuclear Piwi does not affect the heterochromatin organization at dual-strand clusters (Klenov et al., 2014; Rangan et al., 2011; Sienski et al., 2015; Yu et al., 2015b). This is likely because clusters are established early in development by maternally deposited piRNAs (Brennecke et al., 2008; Khurana et al., 2011; de Vanssay et al., 2012; Le Thomas et al., 2014; Akkouche et al., 2017). Altogether, similar to siRNA-mediated transcriptional gene silencing in yeast and plants, in flies, Piwi proteins bound with piRNAs transcriptionally repress their targets through heterochromatin assembly.

While Piwi is required for establishing heterochromatin at many transposons, in *piwi* mutants, some transposons are expressed despite little change in H3K9 methylation or HP1a enrichment (Klenov et al., 2011; Sienski et al., 2012; Klenov et al., 2014). This suggests that Piwi can repress transposon transcription independent of heterochromatin formation (Klenov et al., 2014). Furthermore, heterochromatin assembly and transcriptional repression may be parallel pathways. One protein that is required for transposon repression but is not required for heterochromatin formation is the enigmatic Maelstrom (Mael) (Lim and Kai, 2007; Sienski et al., 2012; Muerdter et al., 2013).

Maelstrom is a Mysterious Protein Required for Transposon Repression

mael is a conserved gene that is found in diverse organisms from unicellular protists to humans (Zhang et al., 2008a). The most conserved region, the MAEL domain, includes the six residues: Glu-His-His-Cys-His-Cys (EHHCHC) and has distant similarity to the DnaQ-H 3′ - 5′/DEDD exonuclease family with the RNase H fold (Zhang et al., 2008a). The protist MAEL domain contains the catalytic residues required for nuclease activity and the purified MAEL domain from *Entamoeba histolytica* is able to degrade ssRNA (Zhang et al., 2008a; Chen et al., 2015). In more complex eukaryotes, while the MAEL

domain no longer contains the catalytic residues required for nuclease activity, the EHHCHC residues are still conserved and are thought to confer RNA-binding ability (Zhang et al., 2008a). Consistent with the binding RNA, the MAEL domains of *D. melanogaster*, *Bombyx mori* (silkworm), and *Mus musculus* (mouse) all interact with ssRNA in vitro (Chen et al., 2015; Matsumoto et al., 2015). In addition to the MAEL domain, the fly Mael protein has a partial potential HMG domain while the mouse MAEL protein contains a more intact HMG-domain (Findley, 2003; Zhang et al., 2008a). HMG proteins are a diverse group of proteins that reversibly bind DNA and are involved with many cellular functions (Travers, 2000; Bianchi and Agresti, 2005). Thus, Mael is an intriguing protein that could potentially interact with both RNA and DNA.

The initial hint that Mael was involved with the piRNA pathway was its localization to the nuage (Findley, 2003). Furthermore, localization of Mael to the nuage was dependent on piRNA pathway genes like *aub* and *vasa* (Findley, 2003; Lim and Kai, 2007). Although Mael is enriched at the perinuclear nuage, Mael is also found dispersed in the cytoplasm and in the nucleus as well (Findley, 2003; Sienski et al., 2012). Interestingly, Mael transits between the nucleus and the cytoplasm (Findley, 2003). A point mutation to the MAEL domain prevents Mael from accumulating in the nucleus and the nuage while the loss of the partial HMG domain leads to a mild increase in Mael in the nucleus (Sienski et al., 2012). Therefore, while Mael is found in many cellular compartments, its localization is dependent on its intact domains.

mael was discovered in a similar manner to many other piRNA pathway genes and was named for the distinct pattern in which the mutant oocytes rapidly mix their cytoplasmic contents, called “cytoplasmic streaming” (Theurkauf et al., 1992; Clegg et al., 1997; Quinlan, 2016). Furthermore, female *mael* mutant flies were sterile and their oocytes had posterior patterning defects and mislocalized microtubule-organizing centers (MTOCs; Clegg et al., 1997; Clegg et al., 2001; Findley, 2003; Cook et al., 2004; Sato et al., 2011). One explanation for the sterility and the patterning defects

seen in *mael* mutants is the loss of transposon repression (Lim and Kai, 2007; Sienski et al., 2012; Czech et al., 2013; Muerdter et al., 2013). Moreover, transposon repression and fertility require intact MAEL and HMG domains (Sienski et al., 2012).

Although it is still unknown how Mael represses transposons, it is thought to function downstream of Piwi because Mael does not appear to be required for heterochromatin formation or piRNA production (Sienski et al., 2012; Muerdter et al., 2013). Much of what we know about Mael was discovered in ovarian somatic cells (OSCs), a cell line that resembles somatic follicle cells (McCarty et al., 2009; Lau et al., 2009; Saito et al., 2009). OSCs express many of the same piRNA pathway proteins also expressed in somatic follicle cells such as Piwi and Mael (Saito et al., 2009; Sienski et al., 2012). Furthermore, like in the ovaries, both Piwi and Mael are required to repress transposons in OSCs (Sienski et al., 2012; Muerdter et al., 2013). In OSCs, Piwi is required for depositing H3K9me3 at euchromatic insertions while depletion of Mael led to a minor decrease but slight spread of H3K9me3 at the same euchromatic transposon insertions (Sienski et al., 2012). Moreover, neither tethering Mael to a reporter transcript or upstream of the TSS of another reporter resulted in silencing (Sienski et al., 2015). While these experiments suggest that merely the presence of Mael alone is not sufficient to repress transcription, more work is needed to understand the mechanism of Mael-mediated transcription repression. Finally, the loss of Mael did not appear to have a major effect on piRNAs in OSCs or ovaries (Sienski et al., 2012; Muerdter et al., 2013). Taken together, evidence suggests that Mael is a downstream effector of Piwi that represses transposons independent of heterochromatin formation.

In *mael* mutant ovaries, both the increase in transposon expression and defects in pattern specification are consistent with activation of the DNA damage signaling pathway. However, mutation of an essential DNA damage signaling gene, *mnk*, did not rescue the patterning defect in *mael* mutant oocytes (Klattenhoff et al., 2007; Sato et al., 2011). Therefore, Mael likely has functions outside the piRNA pathway and may be

directly involved with axis specification through its interactions with the microtubule-organizing center (Sato et al., 2011). Furthermore, Mael may also play a role in promoting oocyte determination and preventing stem cell differentiation (Pek et al., 2009; Pek et al., 2012). While Mael is not required for establishing heterochromatin at transposons, Mael may prevent stem cell differentiation by transcriptionally repressing *miR-7* by accumulating H3K9me3 and HP1a (Pek et al., 2009).

In mice, MAEL is also required for fertility and repressing transposons in the germline (Costa et al., 2006; Soper et al., 2008; Aravin et al., 2009; Castaneda et al., 2014). Furthermore, in both flies and mice, Mael is expressed in all stages of germ cell development (Costa et al., 2006; Soper et al., 2008; Aravin et al., 2009; Dufourt et al., 2013), however, mouse MAEL expression is also regulated by the transcription factor A-MYB, which is required for initiating production of pachytene piRNAs and transcription of many other piRNA pathway genes (Li et al., 2013). Although MAEL is also localized to the perinuclear nuage in mouse testes, in fetal gonocytes MAEL is found in a subset of germinal granules called “piP-bodies” (Soper et al., 2008; Aravin et al., 2009). Along with MAEL, piP-bodies contain MIWI2 and TDRD9 from the piRNA pathway as well as typical components of P-bodies like GW182, DCP1a, DDX6/p54, and XRN1 (Aravin et al., 2009). In mice, MIWI2 is the only PIWI protein that localizes to the nucleus and is required for DNA methylation of transposons (Carmell et al., 2007; Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008). However, in both flies and mice, Mael does not appear to play a major role in heterochromatin formation at transposons (Aravin et al., 2009; Sienski et al., 2012).

Since its discovery, many people have been pulled into studying the mysterious protein Maelstrom and its role in the piRNA pathway. The pleiotropic phenotypes observed in *mael* mutants have only muddied the waters, but through these studies, Mael has been implicated in diverse cellular functions from cytoskeleton organization to heterochromatin formation. Recent studies have provided insight into how Mael

functions, but many questions still remain. My work encompasses experiments ranging from relatively simple genetic screens using one of Muller's original PEV fly lines to measuring nascent transcription using deep sequencing technology and computational analyses. This thesis hopes to clarify how Mael fits into the piRNA pathway to repress transposons.

CHAPTER II: Maelstrom Represses Canonical RNA Polymerase II Transcription in Dual-Strand piRNA Clusters

DISCLAIMER

This work was the joint effort among the authors: Timothy H. Chang (THC), Eugenio Mattei (EM) and Phillip D. Zamore (PDZ). THC conducted genetic experiments and constructed the sequencing libraries. EM performed computational analyses of sequencing data. PDZ supervised the project. All authors provided critical review of the data and manuscript.

SUMMARY

In *Drosophila*, 23–30 nt long PIWI-interacting RNAs (piRNAs) direct the protein Piwi to silence germline transposons. Most germline piRNAs derive from heterochromatic transposon graveyards that are transcribed from both genomic strands: dual-strand piRNA clusters. These piRNA sources are marked by the Heterochromatin Protein 1 homolog, Rhino (Rhi), which facilitates their promoter-independent transcription, suppresses splicing, and inhibits transcription termination. Here, we report that DEDD family nuclease-like protein Maelstrom (Mael), represses canonical, promoter-dependent transcription in dual-strand clusters, allowing Rhi to initiate piRNA precursor transcription. In addition to Mael, the piRNA biogenesis factor, Armitage (Armi), and Piwi, but not Rhi are also required to repress canonical transcription in dual-strand clusters. We propose that Armi, Piwi, and Mael work in a pathway to repress transcription of individual transposons within clusters, while Rhi allows non-canonical transcription of the clusters into piRNA precursors.

PREFACE

In *Drosophila melanogaster*, 23–30 nt long PIWI-interacting RNAs (piRNAs) direct transposon silencing by serving as guides for Argonaute3 (Ago3), Aubergine (Aub), and Piwi, the three fly PIWI proteins (Aravin et al., 2001; Girard et al., 2006; Aravin et al., 2006; Grivna et al., 2006b; Lau et al., 2006; Vagin et al., 2006). In the germ cell cytoplasm, Aub and Ago3 increase the abundance their guide piRNAs via the ping-pong cycle, a feed-forward amplification loop in which cycles of piRNA-directed cleavage of sense and antisense transposon-derived long RNAs generate new copies of the original piRNAs—secondary piRNAs—in response to transposon transcription (Brennecke et al., 2007; Gunawardane et al., 2007). In addition to amplifying piRNAs, the ping-pong pathway also produces long 5′ monophosphorylated RNA that enter the primary piRNA pathway, which uses the long RNA to generate head-to-tail strings of piRNAs bound to Piwi, and to a lesser extent, Aub (Han et al., 2015; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015). Unlike Ago3 and Aub, Piwi acts in both the germline and the adjacent somatic follicle cells and represses transposon transcription rather than cleaving their transcripts (Cox et al., 2000; Brennecke et al., 2007; Malone et al., 2009; Klenov et al., 2011). Nuclear Piwi is believed to bind nascent RNA transcripts, and, by binding the protein Panoramix, tethers the histone methyltransferase SETDB1 to transposon-containing loci. SETDB1 trimethylates histone H3 on lysine 9 (H3K9me3), a histone modification required to create repressive constitutive heterochromatin (H3K9; Rangan et al., 2011; Sienski et al., 2012; Donertas et al., 2013; Le Thomas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013; Rozhkov et al., 2013; Klenov et al., 2014; Sienski et al., 2015; Yu et al., 2015b; Aravin et al., 2008; Shpiz et al., 2011; Post et al., 2014).

piRNA precursor RNAs are transcribed from heterochromatic loci called piRNA clusters. piRNA clusters can span >100 kbp and comprise transposons and transposon

fragments that record a species' evolutionary history of transposon invasion (Brennecke et al., 2007; Lagarrigue et al., 2013; Aravin et al., 2008; Fu et al., 2018b). *D. melanogaster* piRNA clusters can be uni-strand, transcribed from one genomic strand, or dual-strand, transcribed from both genomic strands. Uni-strand clusters, such as the ~180 kbp *flamenco* (*flam*) locus, silence transposons in somatic follicle cells (Pelisson et al., 1994; Prud'homme et al., 1995; Robert et al., 2001; Sarot et al., 2004; Mevel-Ninio et al., 2007; Pelisson et al., 2007), while dual-strand clusters, such as the ~250 kbp *42AB* locus, predominate in the germline (Malone et al., 2009). Some uni-strand clusters, e.g., *cluster2*, are active in both tissues.

Standard promoter-initiated, Pol II transcription generates spliced, polyadenylated precursor piRNAs from *flam* (Robert et al., 2001; Mevel-Ninio et al., 2007; Goriaux et al., 2014). In contrast, dual-strand clusters generally lack conserved promoters. Instead, the Heterochromatin Protein 1 homolog Rhino (Rhi) binds to the H3K9me3 present on the piRNA clusters, to which it can tether additional proteins (Lachner et al., 2001; Cogoni and Macino, 1999; Klattenhoff et al., 2009; Zhang et al., 2012; Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014; Yu et al., 2015a). One Rhi-associated protein, Moonshiner (Moon), is a germline-specific TFIIA-L paralog that allows Pol II to initiate transcription without the TATA-box binding protein, allowing every bound Rhi to be a site of potential transcription initiation (Andersen et al., 2017). Another Rhi-binding protein, Cutoff (Cuff), suppresses splicing and transcription termination (Pane et al., 2011; Mohn et al., 2014; Zhang et al., 2014; Chen et al., 2016). Thus, Rhi promotes "incoherent" transcription, initiating at many sites throughout both strands of a dual-strand cluster, in contrast to the "coherent," promoter-dependent transcription of *flam* and conventional protein-coding genes.

Maelstrom (Mael), a protein with HMG- (Findley, 2003) and MAEL- (Zhang et al., 2008a) domains, has been suggested to play multiple roles in *D. melanogaster* oogenesis and mouse spermatogenesis, including transposon silencing,

heterochromatin formation, and piRNA production. Here, we report that Mael suppresses coherent transcription within dual-strand piRNA clusters. In *mael* mutant ovaries, piRNA clusters produce more transcripts yet heterochromatin organization is largely unaltered. However, without Mael, Piwi or Armitage (Armi), a core piRNA biogenesis protein, transcription initiates from canonical Pol II promoters within dual-strand clusters such as *42AB*. Although Rhi and Cuff are required for incoherent transcription of dual-strand piRNA clusters, they are dispensable for repression of canonical transcription in the clusters. We propose that Mael, Armi, Piwi, and piRNAs collaborate to repress canonical dual-strand cluster transcription, while Rhi serves both to create a transcriptionally permissive chromatin environment and to support incoherent transcription of both DNA strands of dual-strand clusters. Thus, Mael represses promoter-driven transcription of individual, potentially active, transposons embedded within dual strand-clusters, allowing Rhi to transcribe such transposon sequences into piRNA precursors with little potential to be translated into transposon-encoded proteins required for transposition.

RESULTS

Mael Represses Canonical Transcription in Dual-Strand piRNA Clusters

Without Mael, both somatic and germline transposons produce long RNA transcripts (Sienski et al., 2012; Muerdter et al., 2013; Pek et al., 2009; Figure 2.1A). While transposons from many different families increased, overall, transcripts from most protein-coding genes did not change significantly between *mael*$^{M391/r20}$ and control ovaries (Figure 2.1B). In the germline of *mael*$^{M391/r20}$ ovaries, uniquely mapping steady-state RNA abundance from the *gypsy12* long terminal repeat (LTR) transposon increases >20 times ($p = 6.5 \times 10^{-5}$, $n = 3$). Intriguingly, the *gypsy12* RNA derives from the long terminal repeat (LTR) of two *gypsy12* elements, one in the dual-strand piRNA cluster *42AB* (*gypsy12*^{42A14}) and one in *cluster62* (*gypsy12*^{40F7}; Figure 2.2A). The same two *gypsy12* elements are also desilenced in *rhi* and *cuff* mutant ovaries (Zhang et al., 2014). As in *rhi* and *cuff* mutants, RNA from the two *gypsy12* LTRs were spliced and *gypsy12*^{42A14} terminated at a canonical poly(A) site while *gypsy12*^{40F7} initiated at a canonical TATAA sequence in *mael*$^{M391/r20}$ mutants. The increase in steady-state *gypsy12* RNA in *mael*$^{M391/r20}$ mutants reflects an increase in nascent transcription: Global run-on sequencing (GRO-seq; Core et al., 2008) revealed >7.5-fold increase in *mael*$^{M391/r20}$ mutants ($p = 3.0 \times 10^{-4}$, $n = 3$; Figure 2.2B). Moreover, lysine 4 trimethylatation of histone H3 (H3K4me3), a chromatin mark associated with active, promoter-driven transcription (Bernstein et al., 2002; Santos-Rosa et al., 2002; Schneider et al., 2004), is ≥3 times higher across the two desilenced *gypsy12* sequences in *mael*$^{M391/r20}$ ovaries ($n = 2$; Figure 2.2C). These data suggest that in the absence of Mael, Pol II initiates transcription from a promoter residing within the *gypsy12* LTR.

Figure 2.1

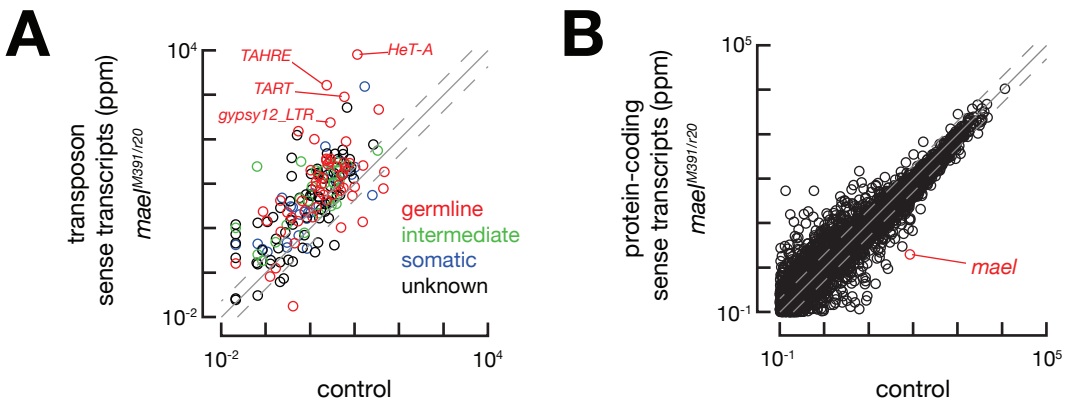

A
transposon sense transcripts (ppm)
mael^M391/r20
10^4
10^-2
TAHRE
HeT-A
TART
gypsy12_LTR
germline
intermediate
somatic
unknown
10^-2
10^4
control
B
protein-coding sense transcripts (ppm)
mael^M391/r20
10^5
10^-1
mael
10^-1
10^5
control

Figure 2.1: Many transposons are overexpressed without Mael

(A) Scatter plots comparing sense RNA-seq reads from *mael*$^{M391/r20}$ and control ovaries that uniquely map to germline (red), intermediate (green), somatic (blue), or unknown (black) transposons. (B) Scatter plots comparing sense RNA-seq reads from *mael*$^{M391/r20}$ and control ovaries that uniquely map to protein-coding genes. Transcripts that map to *mael* are depicted with a red circle. The hashed grey line signifies a ≥2-fold change. Data are the mean of three biological replicates.

Figure 2.2

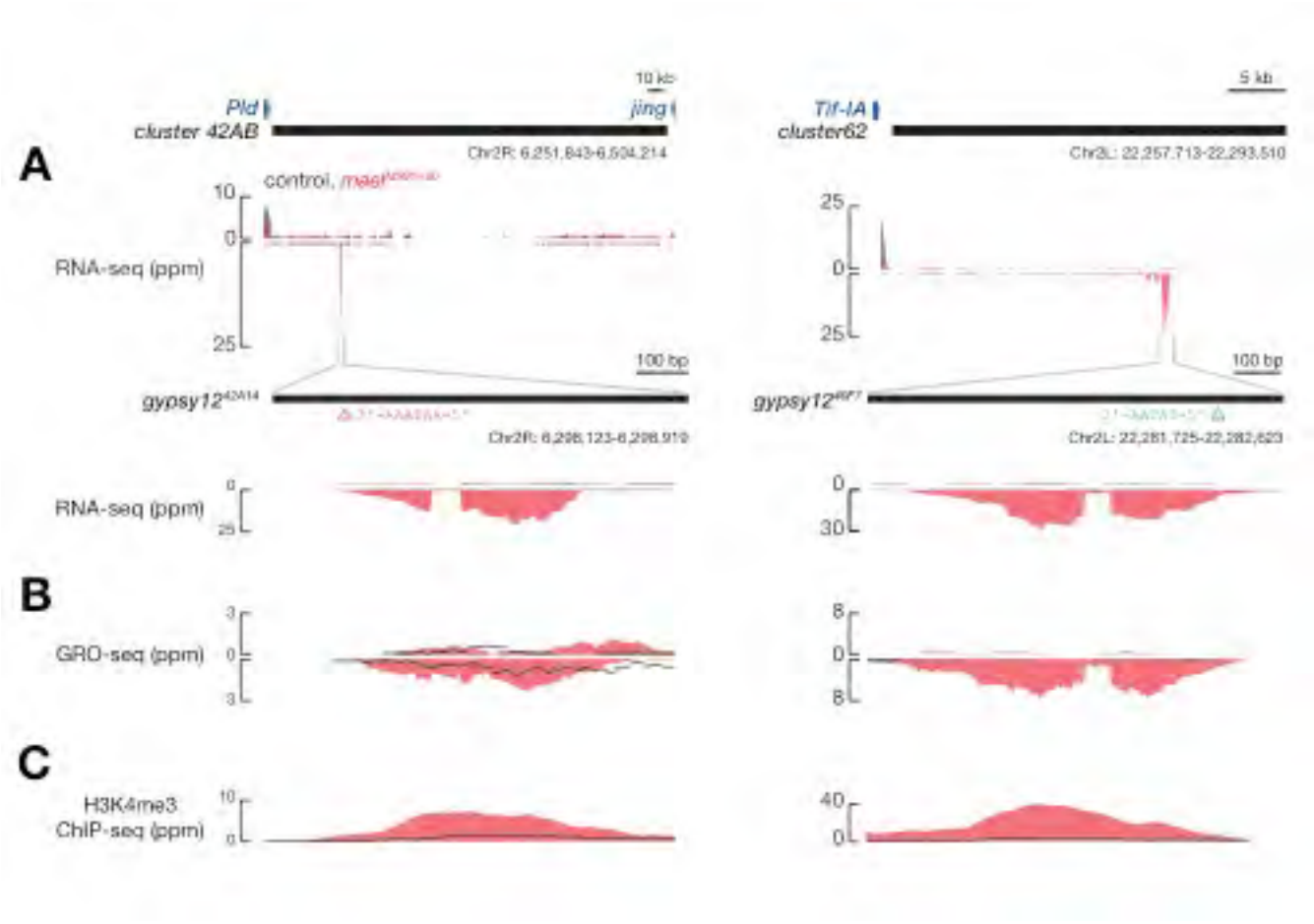

A
B
C
Pld
cluster 42AB
jing
10 kb
Chr2R: 6,251,843-6,504,214
control, mael
RNA-seq (ppm)
100 bp
gypsy12
Chr2R: 6,298,123-6,298,919
GRO-seq (ppm)
H3K4me3
ChIP-seq (ppm)
Tlf-1A
cluster62
5 kb
Chr2L: 22,257,713-22,293,510
gypsy12
Chr2L: 22,281,725-22,282,623

Figure 2.2: Two *gypsy12* LTRs in dual-strand piRNA clusters are active in *mael*$^{M391/r20}$ ovaries

(A) RNA-seq profiles for *42AB* (left) and *cluster62* (right). A higher resolution expansion of the two *gypsy12* LTRs, *gypsy12*^{42A14} and *gypsy12*^{40F7}, located in *42AB* and *cluster62*, respectively, is shown below displaying (A) RNA-, (B) GRO-, and (C) H3K4me3 ChIP-seq profiles. TSSs and TTSs are shown with green and red triangles, respectively. Reads from *mael*$^{M391/r20}$ mutants are shown in red while wild-type controls are shown in black. All signals are uniquely mapping and displayed as ppm (parts per million genome mappers).

Loss of Mael also led to increased use of the canonical Pol II promoters flanking the unusual dual-strand *cluster38C1*. Unlike typical piRNA clusters, *cluster38C1* can sustain piRNA precursor production in mutants that disrupt Rhi-dependent incoherent transcription (Mohn et al., 2014; Chen et al., 2016; Andersen et al., 2017). In $mael^{M391/r20}$ ovaries, transcription initiated at canonical TATA-box sequences flanking the cluster but did not terminate at canonical poly(A) signal sequences ~400 bp downstream of either flanking promoter (Figure 2.3A). Furthermore, steady-state abundance of uniquely mapping *cluster38C1* RNA >150 nt long increased ~15-fold in $mael^{M391/r20}$ ovaries ($p = 3.1 \times 10^{-4}$). Like $gypsy12^{42A14}$ and $gypsy12^{40F7}$, we detected an increase in nascent transcription from both the plus (*mael*/control = 3 ± 1-fold; $p = 0.046$) and minus (*mael*/control = 5 ± 2; $p = 0.014$) genomic strands (Figure 2.3B). Unlike the two *gypsy12* LTRs, however, we did not detect a change in the active chromatin mark H3K4me3 at *cluster38C1* in $mael^{M391/r20}$ ovaries (Figure 2.3C). A combination of canonical and Rhi-dependent incoherent transcription produces piRNA precursors from *cluster38C1*. We speculate that because the wild-type level of H3K4me3 at *cluster38C1* already suffices to initiate transcription at the flanking promoters, no further increase occurs in $mael^{M391/r20}$ mutants.

Figure 2.3

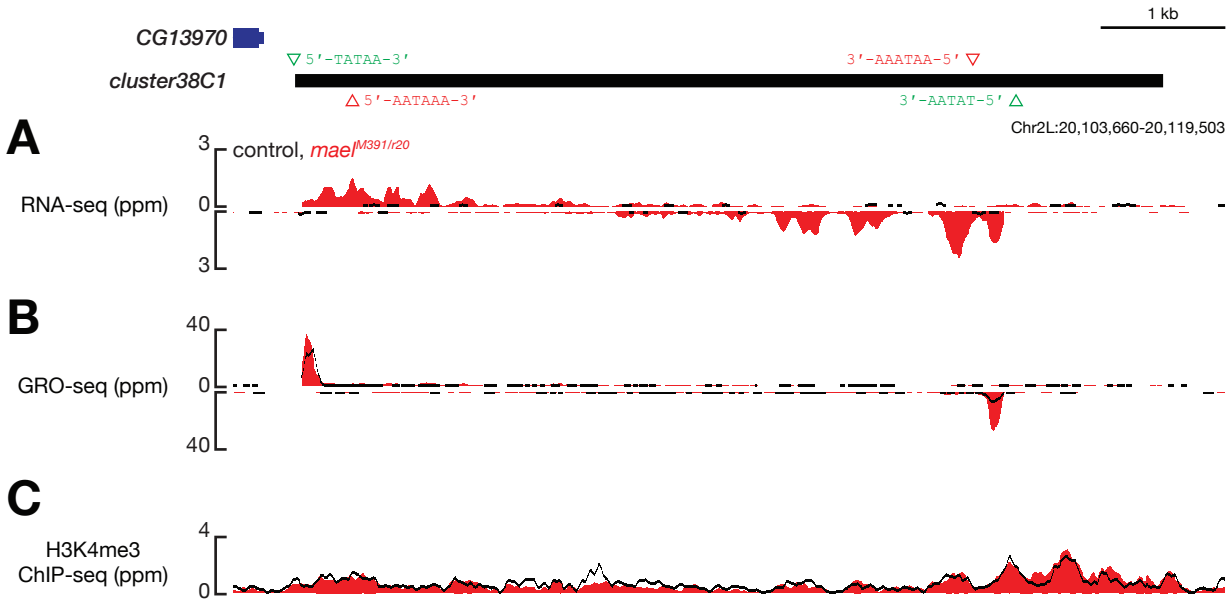

1 kb
CG13970
cluster38C1
5′-TATAA-3′
3′-AAATAA-5′
5′-AATAAA-3′
3′-AATAT-5′
Chr2L:20,103,660-20,119,503
A
control, mael M391/r20
RNA-seq (ppm)
3
0
3
B
GRO-seq (ppm)
40
0
40
C
H3K4me3
ChIP-seq (ppm)
4
0

Figure 2.3: Increased transcription of *cluster38C1* in *mael*$^{M391/r20}$ ovaries

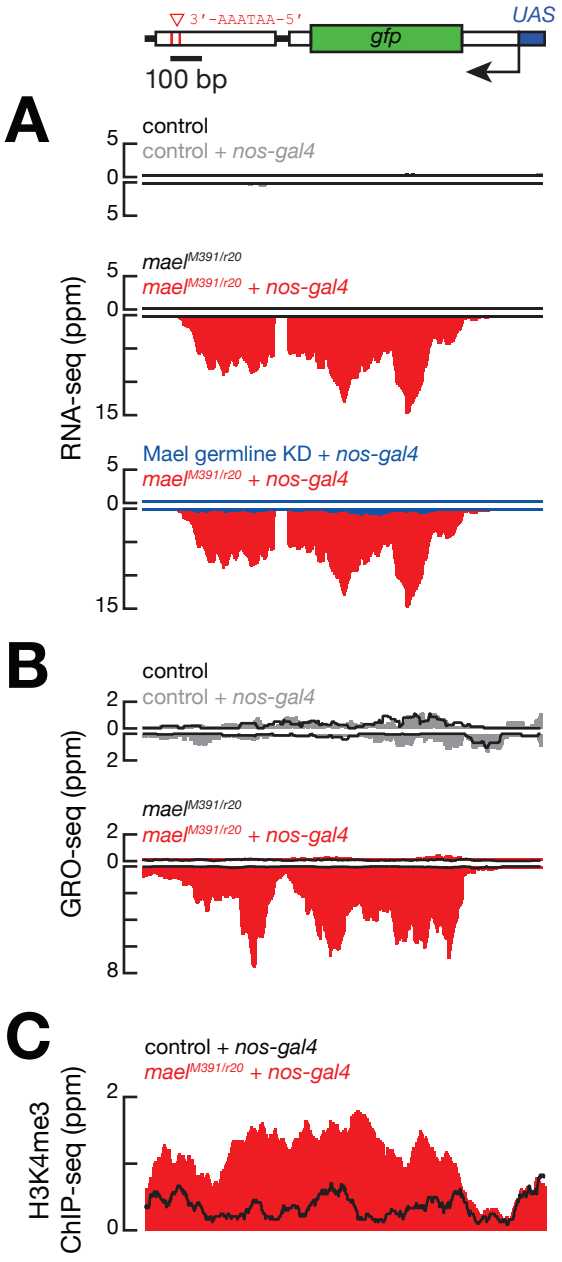
(A) RNA-, (B) GRO-, and (C) H3K4me3 ChIP-seq profiles for *cluster38C1*. Flanking promoters with canonical TATA sequences (green triangle) initiate transcription of *cluster38C1* in both *mael*$^{M391/r20}$ (red) and control ovaries (black). Canonical poly(A) sequences are shown and marked by a red triangle.

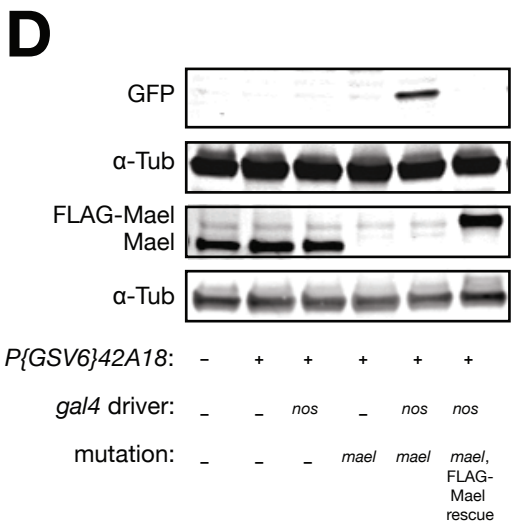
A Reporter for Transcription in Dual-Strand piRNA Clusters

piRNA clusters are highly repetitive, complicating bioinformatics analysis. The *P{GSV6}42A18* fly strain inserts an inducible *UASt-GFP* gene with an intron in the 3′ untranslated region (UTR) into *42AB*, allowing this transgene to be used as a proxy for canonical euchromatic transcription within piRNA clusters (Chendrimada et al., 2005). Like *42AB* itself, *P{GSV6}42A18* requires Rhi, Armi, and Piwi to produce sense and antisense piRNAs (Han et al., 2015). The *P{GSV6}42A18* transgene resembles the piRNA cluster in which it resides, with a high density of H3K9me3, HP1a, and Rhi across its sequence, and expression of both *gfp* mRNA and protein was essentially undetectable even when the strong transcriptional activator GAL4-VP16 was co-expressed from the germline-specific *nanos* promoter (Figures 2.4A, 2.4B, and 2.4D).

In $mael^{M391/r20}$ ovaries, the *P{GSV6}42A18* transgene, driven by GAL4-VP16, produced correctly spliced *gfp* mRNA that terminated at a canonical polyadenylation signal sequence (Figure 2.4A); the appearance of *gfp* mRNA was accompanied by increased transcription (Figure 2.4B) and H3K4me3 across the *gfp* transgene (Figure 2.4C). Moreover, the *gfp* mRNA in $mael^{M391/r20}$ mutants was translated into full-length GFP protein (Figure 2.4D). A transgene encoding FLAG-Mael restored repression of GFP in $mael^{M391/r20}$ ovaries demonstrating loss of Mael, and not a secondary mutation, caused inappropriate GFP expression from the transgene inserted into *42AB* (Figure 2.4D). Interestingly, germline knockdown of Mael, which depletes Mael in adult ovaries but does not affect maternally deposited Mael, did not cause the derepression of *gfp* (Figure 2.4A). This suggests that Mael is required in early embryo development to silence transcription. Together, our data suggest that Mael represses canonical, promoter driven Pol II transcription in dual-strand piRNA clusters.

Figure 2.4





	mutation	*gal4* driver	GFP Western signal	*p*-value
P{GSV6}42A18	none	none	3 ± 2	0.097
	none	none	9 ± 4	
	none	*nos*	19 ± 6	0.057
	mael$^{M391/r20}$	none	30 ± 10	0.085
	mael$^{M391/r20}$	*nos*	400 ± 100	0.014
	mael$^{M391/r20}$, Mael Rescue	*nos*	40 ± 20	0.19

Figure 2.4: Mael represses the canonical transcription of *P{GSV6}42A18*

(A) A schematic showing *P{GSV6}42A18*, a transgene inserted into *42AB*. *P{GSV6}42A18* contains an inducible *UASt-GFP* gene with an intron and canonical poly(A) sites in the 3′ UTR. RNA-seq profiles for *P{GSV6}42A18* from $mael^{M391/r20}$ and control ovaries, with or without *nanos-gal4*, and Mael germline knockdown with *nanos-gal4*. (B) GRO-seq profiles from $mael^{M391/r20}$ and control ovaries with or without *nanos-gal4*. (C) H3K4me3 ChIP-seq profiles displaying $mael^{M391/r20}$ (red) and control (black) ovaries expressing *nanos-gal4*. (D) Western blots for GFP, Mael, or α-Tubulin (α-Tub) from ovaries with the genotype given below. GFP Western signal is the average of three biological replicates normalized to α-Tub and is given in arbitrary units. *p*-values were measured using an unpaired, two-tailed t-test compared to w^{1118}*; P{GSV6}42A18*/+; +.

Many Pol II Promoters are Activated in *mael* Mutant Ovaries

Without Mael, RNA accumulates from both individual euchromatic transposons outside of clusters (Sienski et al., 2012; Muerdter et al., 2013) and heterochromatic transposon sequences within clusters (Figure 2.5A). To further test the idea that Mael represses canonical transcription at sites of Rhi-driven incoherent transcription, we examined in more detail those transposons whose steady-state RNA abundance increased in *mael*$^{M391/r20}$ ovaries. Overall, uniquely mapping steady-state RNA from 1485 individual transposon loci (410 in and 1075 out clusters) increased ≥2-fold (FDR ≤ 0.05; n = 6) in *mael*$^{M391/r20}$ ovaries (Figure 2.5B). Among these derepressed transposons, 182 transposon loci overlapped with H3K4me3 peaks that more than doubled in *mael* mutant ovaries compared to control (70 in and 112 out clusters; Figure 2.5C**)**. Moreover, spliced transcripts—measured by the abundance of unambiguously mapping exon-exon junction RNA-seq reads—more than doubled for 29 (13 in and 16 out clusters) of these182 transposon loci in *mael*$^{M391/r20}$ ovaries (Figure 2.5D).

Figure 2.5

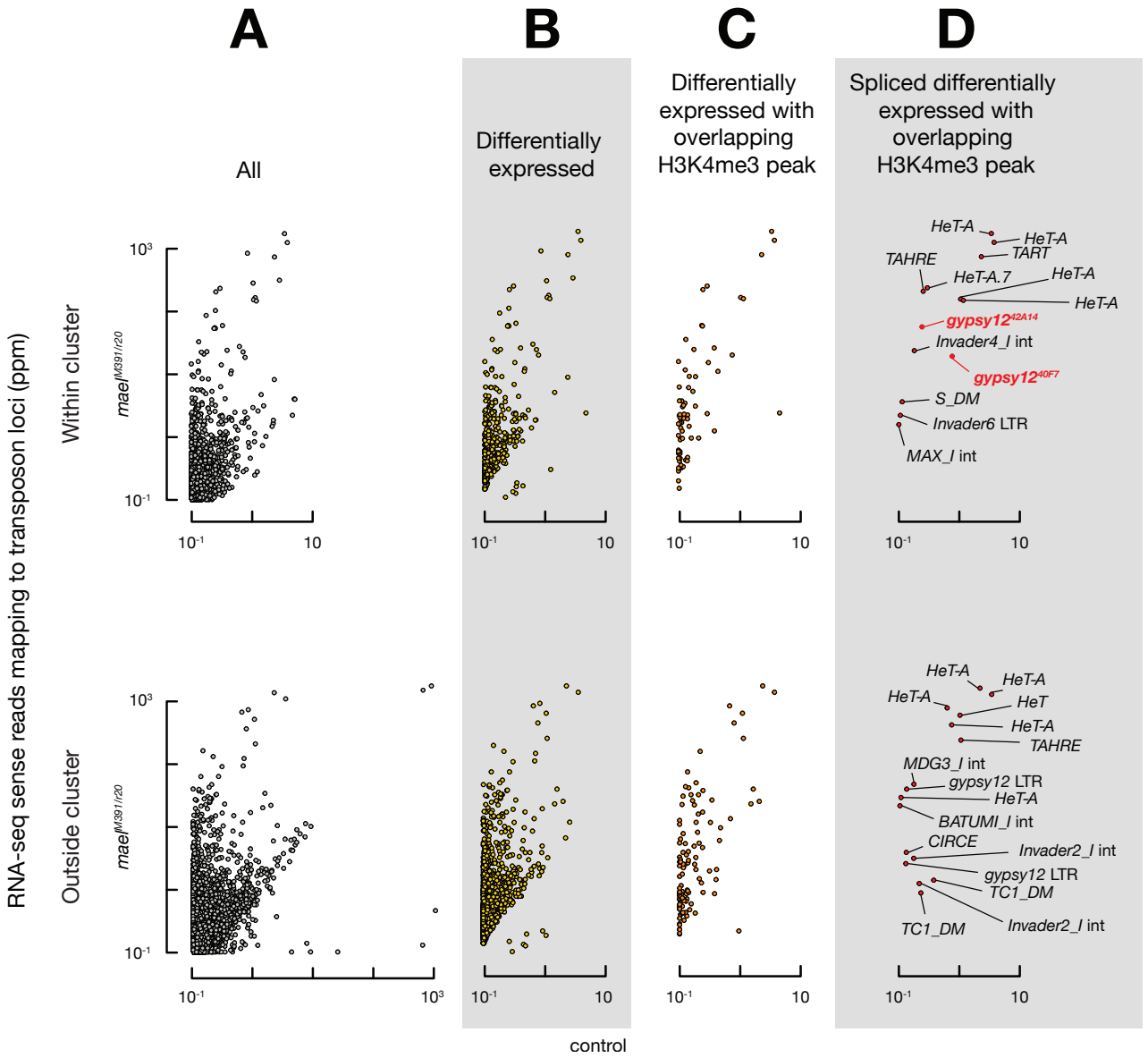

A
B
C
D
All
Differentially expressed
Differentially expressed with overlapping H3K4me3 peak
Spliced differentially expressed with overlapping H3K4me3 peak
RNA-seq sense reads mapping to transposon loci (ppm)
Within cluster
Outside cluster
mael^M391/r20
control
HeT-A
TAHRE
TART
HeT-A.7
gypsy12^42A14
Invader4_I int
gypsy12^40F7
S_DM
Invader6 LTR
MAX_I int
HeT
MDG3_I int
gypsy12 LTR
BATUMI_I int
CIRCE
Invader2_I int
TC1_DM

Figure 2.5: Active transposon promoters are canonically transcribed in *mael*$^{M391/r20}$ ovaries

(A) Scatter plots comparing sense RNA-seq reads from *mael*$^{M391/r20}$ and control ovaries (n = 6) that uniquely mapped to 11,810 individual transposon loci inside and outside piRNA clusters. (B) RNA-seq reads that were differentially expressed (≥2-fold increase; FDR ≤ 0.05) between *mael*$^{M391/r20}$ and control ovaries were kept (1485 total transposons; 410 within and 1075 outside clusters). (C) Using a set of consensus H3K4me3 peaks from *mael*$^{M391/r20}$ and control ovaries, 182 transposon loci (70 within and 112 outside clusters) also had overlapping H3K4me3 peaks that had more signal in *mael*$^{M391/r20}$ mutants (≥2-fold increase). (D) Scatter plots comparing differentially expressed spliced transcripts (≥2-fold increase; FDR ≤ 0.05) from *mael*$^{M391/r20}$ and control ovaries that also had overlapping H3K4me3 peaks (13 within and 16 outside clusters). *gypsy12*^{42A14} and *gypsy12*^{40F7} are shown in red.

In flies, most promoters are uni-directional, therefore antisense reads are unlikely to be products of canonical promoter-dependent transcription (Nechaev et al., 2010). Using the previous strategy of analysis on antisense rather than sense transcripts (Figure 2.6A), in *mael*$^{M391/r20}$ mutants, antisense transposon transcripts within and without piRNA clusters also increased (153 within and 274 outside clusters; Figure 2.6B), had overlapping H3K4me3 peaks (19 within and 20 outside clusters; Figure 2.6C), and were spliced (6 within and 7 outside clusters; Figure 2.6D). Although sense transposon transcripts were expressed ~10-fold higher than antisense, the more than 2-fold increase in spliced antisense transcripts with overlapping H3K4me3 peaks suggests that cryptic antisense promoters may be active in *mael*$^{M391/r20}$ ovaries. We conclude that loss of Mael increases canonical Pol II transcription from both euchromatic transposons outside piRNA clusters and from heterochromatic transposons inside piRNA clusters.

Figure 2.6

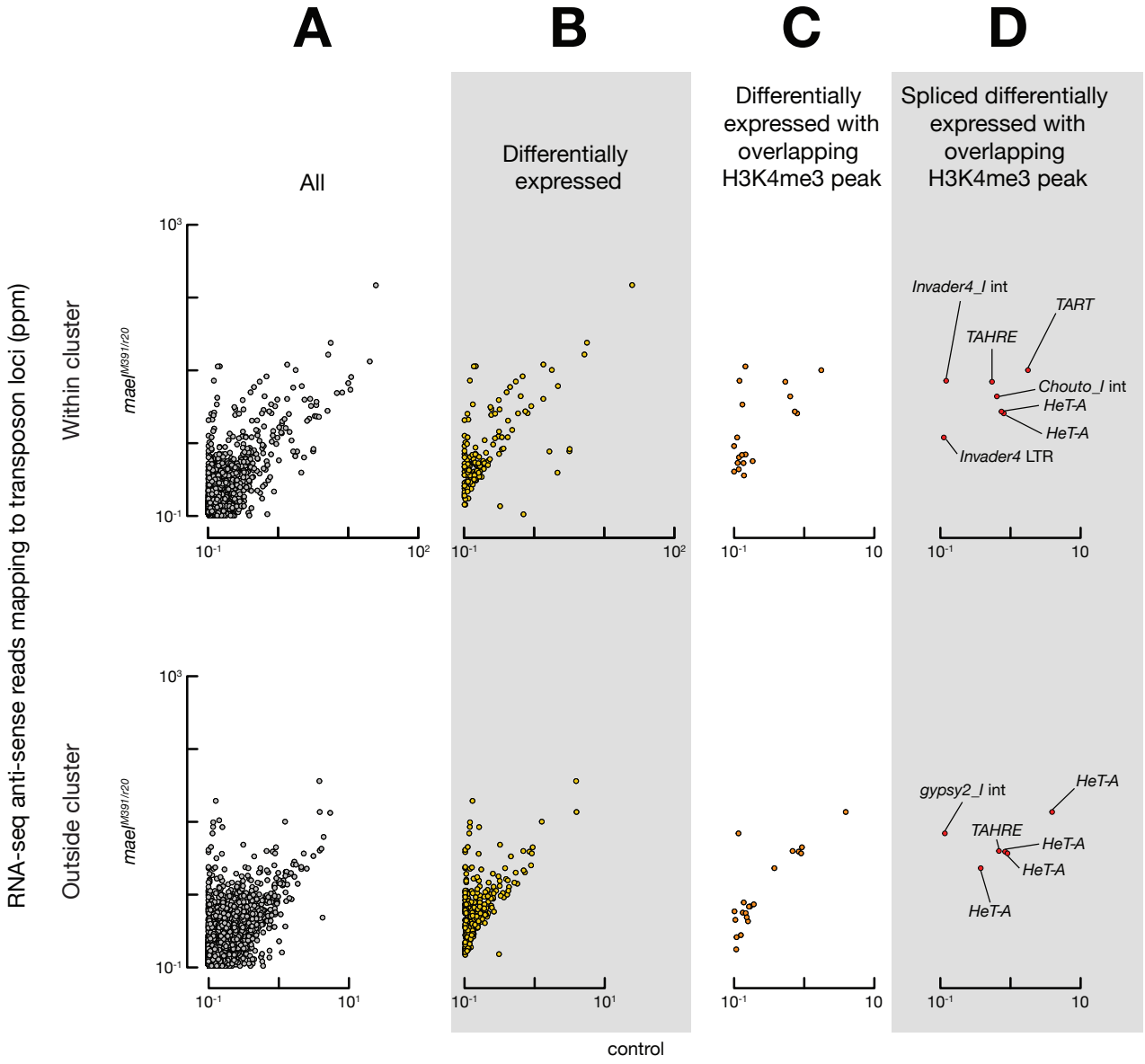

A
B
C
D
All
Differentially expressed
Differentially expressed with overlapping H3K4me3 peak
Spliced differentially expressed with overlapping H3K4me3 peak
RNA-seq anti-sense reads mapping to transposon loci (ppm)
Within cluster
Outside cluster
mael^M391/20
control
Invader4_I int
TART
TAHRE
Chouto_I int
HeT-A
HeT-A
Invader4 LTR
gypsy2_I int
HeT-A
TAHRE
HeT-A
HeT-A
HeT-A

Figure 2.6: Potential cryptic promoters in transposons are active in *mael*$^{M391/r20}$ mutant ovaries

(A) The same analyses used for sense RNA-seq reads in Figure 2.4 was also used for antisense reads. Scatter plots comparing antisense RNA-seq reads from *mael*$^{M391/r20}$ and control ovaries (n = 6) that uniquely mapped to 10,745 individual transposon loci inside and outside piRNA clusters. (B) Out of these individual loci, 427 transposons (153 within and 274 outside clusters) had differentially expressed antisense transcripts (≥2-fold increase; FDR ≤ 0.05) in *mael*$^{M391/r20}$ ovaries. (C) Without Mael, 39 transposon loci (19 within and 20 outside clusters) also had overlapping H3K4me3 peaks that had a ≥2-fold increase. (D) 6 individual transposon loci within and 7 outside piRNA clusters fit all the previous criteria and also produced more spliced antisense transcripts (≥2-fold increase; FDR ≤ 0.05).

Heterochromatin is Intact in *mael* Mutant Ovaries

Does loss of heterochromatin at transposons inside and outside piRNA clusters explain the increase in their transcription in *mael*$^{M391/r20}$ ovaries? To test this idea, we used chromatin immunoprecipitation sequencing (ChIP-seq) to examine H3K9me3, HP1a, and Rhi in wild-type control and *mael*$^{M391/r20}$ mutant ovaries. HP1a binds H3K9me3, compacts chromatin, and, like Rhi, decorates piRNA clusters (Bannister et al., 2001; Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002; Vermaak and Malik, 2009; Klenov et al., 2014). Using reads that uniquely mapped to the genome, ovaries with or without Mael had very little change in heterochromatin across piRNA clusters and transposons (Figures 2.7A and 2.7B). One exception, however, were the telomeric transposons, *HeT-A*, *TAHRE*, and *TART*, which had >3-fold decrease in H3K9me3 signal in *mael*$^{M391/r20}$ compared to control ovaries (Figure 2.7B).

We note that one cluster, the sub-telomeric *cluster136*, had a large increase in H3K9me3 and HP1a in *mael*$^{M391/r20}$ mutants (Figure 2.7A). *Cluster136* predominantly consists of one somatic transposon, *gypsy6*. While *gypsy6* is not active in *mael*$^{M391/r20}$ mutant or control ovaries, one possibility for the increase in H3K9me3 and HP1a signal could be a difference in the copy number of *gypsy6* or changes in *cluster136* between the genotypes. Furthermore, while both *mael*^{M391} and *mael*^{r20} strains were outcrossed for five generations, *cluster136* and *mael* are on the same chromosome arm about 4.5 Mb apart.

Figure 2.7

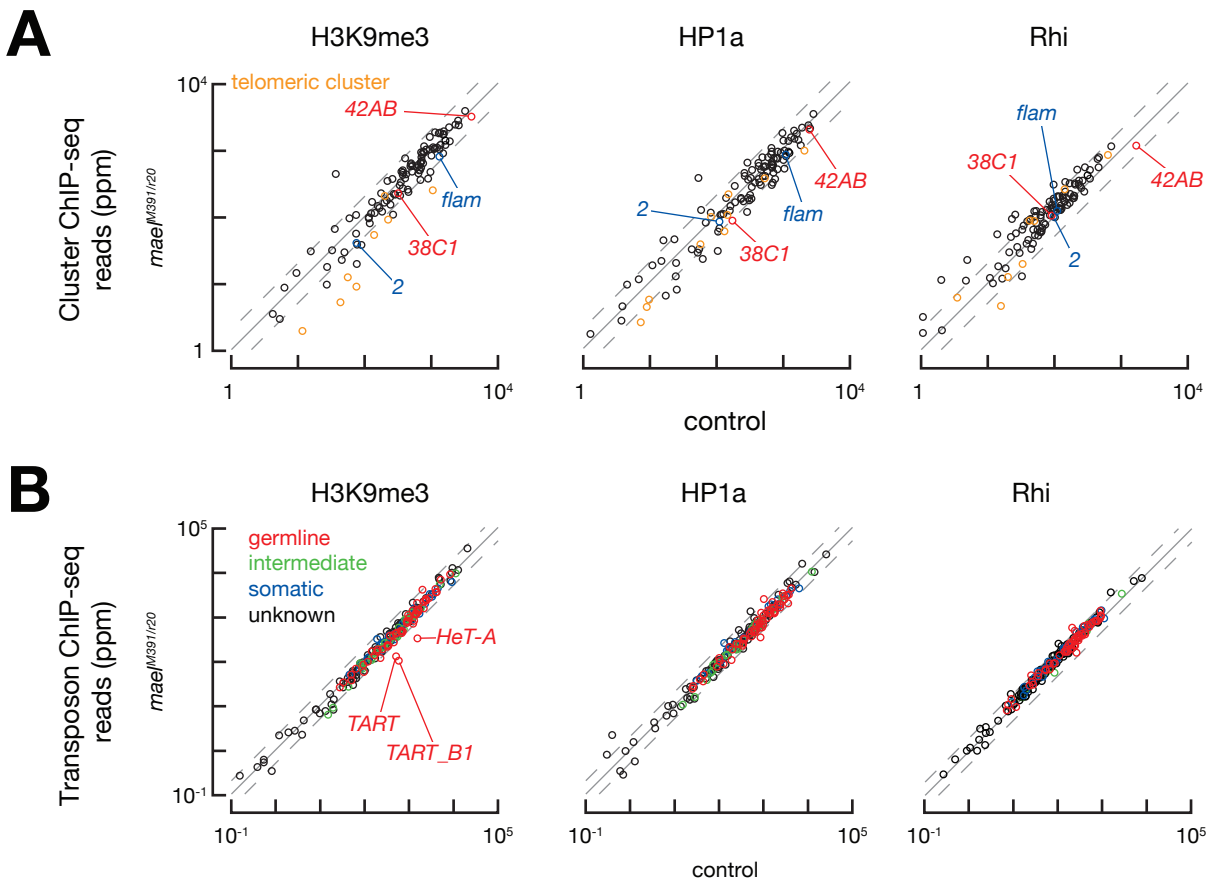

A
H3K9me3
HP1a
Rhi
Cluster ChIP-seq reads (ppm)
mael^M391/r20
telomeric cluster
42AB
flam
38C1
2
control
B
Transposon ChIP-seq reads (ppm)
germline
intermediate
somatic
unknown
HeT-A
TART
TART_B1

Figure 2.7: Mael is not required to establish heterochromatin at piRNA clusters or transposons

(A) Scatter plots on a common y-axis comparing H3K9me3, HP1a, and Rhi ChIP-seq reads that uniquely mapped to piRNA clusters from *mael*$^{M391/r20}$ (y-axis) and control (x-axis) ovaries. Labeled clusters are uni- (blue) or dual-strand (red). Telomeric clusters are displayed as orange circles and the solitary cluster with more H3K9me3 and HP1a in *mael*$^{M391/r20}$ mutants is *cluster136* (see text). (B) Scatter plots on a common y-axis comparing H3K9me3, HP1a, and Rhi ChIP-seq reads that uniquely mapped to transposons from *mael*$^{M391/r20}$ and control ovaries. Transposons are classified as germline-specific (red), soma-specific (blue), intermediate (green), or unknown (black) according to their expression in germline and somatic piRNA pathway mutants (Wang et al., 2015). The hashed grey line signifies a ≥2-fold change. Data are the mean of two biological replicates.

In the two *gypsy12* LTRs found in *42AB* and *cluster62*, *gypsy12*^{42A14} and *gypsy12*^{40F7}, respectively, we observed small changes in H3K9me3, HP1a, and Rhi. There was a slight decrease in Rhi and HP1a in *mael*$^{M391/r20}$ mutants, *gypsy12*^{40F7}, showed a loss of H3K9me3 and HP1a but not Rhi (Figure 2.8). Despite these minor changes in heterochromatin, both *gypsy12* elements also had more H3K4me3 and were still expressed in *mael*$^{M391/r20}$ ovaries (Figure 2.2C). This is consistent with transposon LTRs acting as promoters. In addition, the presence of active and repressive chromatin markers can be found at active genes in heterochromatin, including transposons (Riddle et al., 2011).

Like *gypsy12* in *42AB* and *cluster62*, over the entire *cluster38C1*, we only saw a few minor changes in heterochromatin (Figure 2.9A). Overall, in *mael*$^{M391/r20}$ ovaries, we observed a subtle loss of H3K9me3 and HP1a, while Rhi appeared to increase slightly over the TSS. Finally, despite the canonical transcription of *P{GSV6}42A18*, we did not see any large changes in heterochromatin at this transgene in *mael*$^{M391/r20}$ ovaries (Figure 2.9B). We conclude that the changes in cluster transcription and the loss transposon silencing are unlikely to be due to changes in heterochromatin.

Figure 2.8

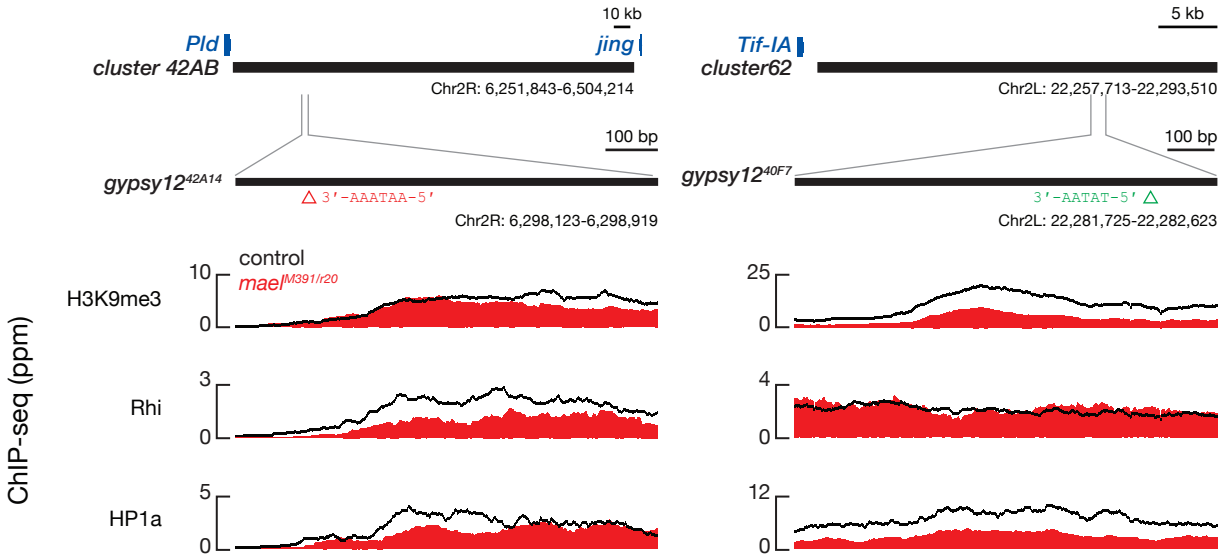

10 kb
Pld
jing
cluster 42AB
Chr2R: 6,251,843-6,504,214
100 bp
gypsy12^42A14
△ 3′-AAATAA-5′
Chr2R: 6,298,123-6,298,919
5 kb
Tif-IA
cluster62
Chr2L: 22,257,713-22,293,510
100 bp
gypsy12^40F7
3′-AATAT-5′ △
Chr2L: 22,281,725-22,282,623
ChIP-seq (ppm)
control
mael^M391/r20
H3K9me3
Rhi
HP1a
10
0
3
0
5
0
25
0
4
0
12
0

Figure 2.8: Minor changes in heterochromatin at *gypsy12*^{42A14} and *gypsy12*^{40F7} in *mael*$^{M391/r20}$ ovaries

H3K9me3, Rhi, and HP1a ChIP-seq profiles for *gypsy12*^{42A14} (left) and *gypsy12*^{40F7} (right) in *mael*$^{M391/r20}$ (red) and control (black) ovaries.

Figure 2.9

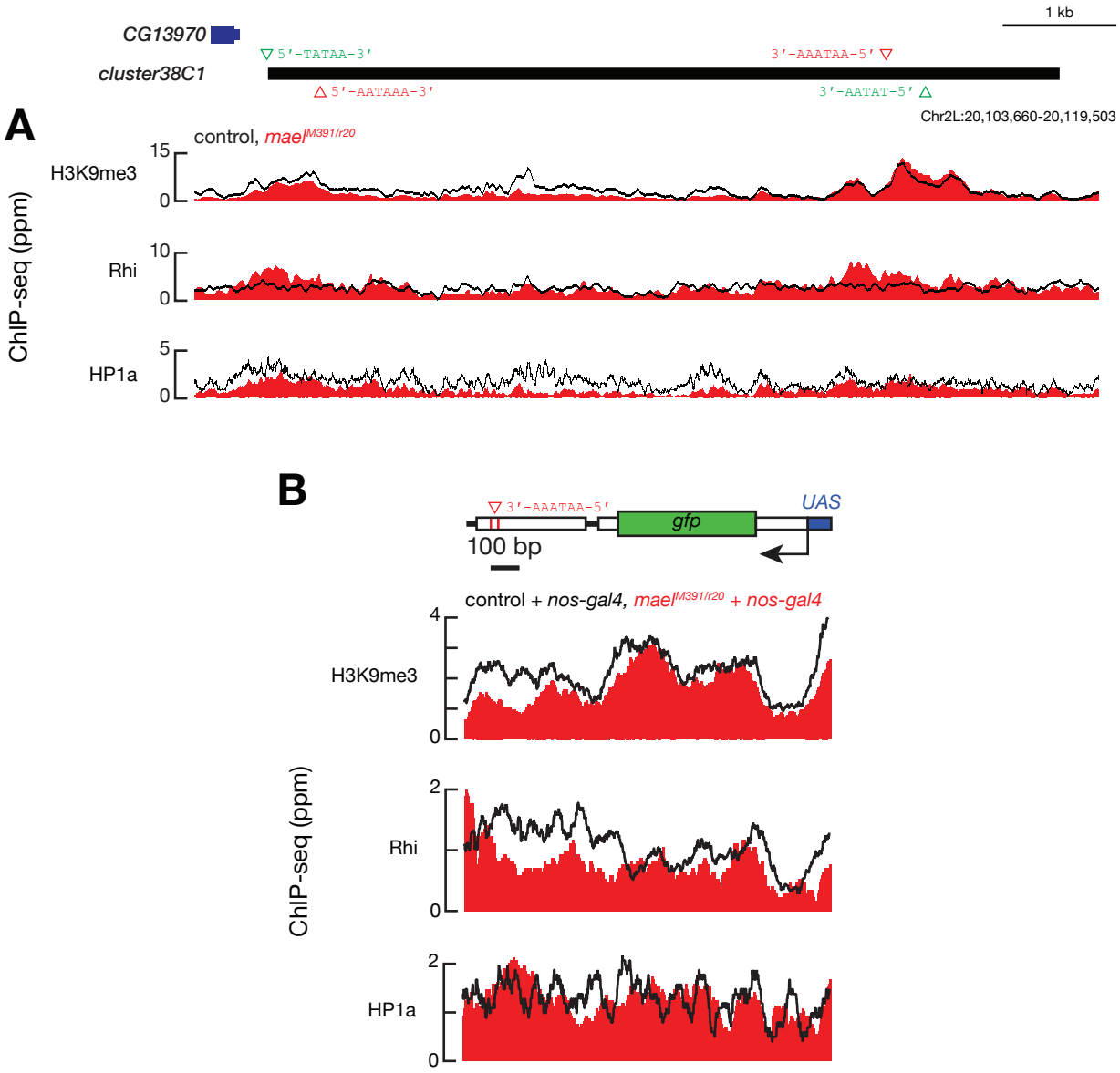

A
1 kb
CG13970
5′-TATAA-3′
3′-AAATAA-5′
cluster38C1
5′-AATAAA-3′
3′-AATAT-5′
Chr2L:20,103,660-20,119,503
control, mael M391/r20
ChIP-seq (ppm)
H3K9me3
Rhi
HP1a
15
0
10
0
5
0
B
3′-AAATAA-5′
UAS
gfp
100 bp
control + nos-gal4, mael M391/r20 + nos-gal4
ChIP-seq (ppm)
H3K9me3
Rhi
HP1a
4
0
2
0
2
0

Figure 2.9: Minor changes in heterochromatin at *cluster38C1* and *P{GSV6}42A18* without Mael

(A) H3K9me3, Rhi, and HP1a ChIP-seq profiles for *cluster38C1* and (B) *P{GSV6}42A18* in *mael*$^{M391/r20}$ (red) and control (black) ovaries.

mael Mutants Produce Fewer piRNAs

Both nascent transcripts and RNA steady-state transcripts from piRNA clusters increase in $mael^{M391/r20}$ mutants, yet these long RNAs produce fewer piRNAs than control ovaries (all piRNAs, *mael*/control = 0.39 ± 0.03, $p = 6.8 \times 10^{-6}$; uniquely mapping piRNAs, *mael*/control = 0.28 ± 0.02, $p = 9.4 \times 10^{-7}$; Figure 2.10A). The abundance of piRNAs from *gypsy12* LTRs (*mael*/control = 0.045 ± 0.003, $p = 5.8 \times 10^{-8}$; Figure 2.10B), dual-strand cluster *cluster38C1* (*mael*/control = 0.10 ± 0.01, $p = 5.9 \times 10^{-7}$), as well as from *P{GSV6}42A18* (*mael*/control = 0.09 ± 0.01, $p = 3.6 \times 10^{-7}$) were all lower in $mael^{M391/r20}$ ovaries (Figure 2.10C). While uni-strand clusters produced less piRNAs in $mael^{M391/r20}$ mutants, dual-strand clusters were more affected. For example, piRNA abundance declined 3–5-fold for the uni-strand clusters *cluster2* (*mael*/control = 0.21 ± 0.02; $p = 8.4 \times 10^{-6}$) and *flam* (*mael*/control = 0.33 ± 0.03; $p = 1.8 \times 10^{-5}$), whereas piRNAs fell 7.6–32-fold for the dual-strand clusters *42AB* (*mael*/control = 0.053 ± 0.005; $p = 9.7 \times 10^{-10}$), *80F* (*mael*/control = 0.13 ± 0.01; $p = 4.6 \times 10^{-7}$), and the telomeric clusters (*mael*/control = 0.031 ± 0.001; $p = 1.0 \times 10^{-8}$; Figure 2.10C)

Figure 2.10

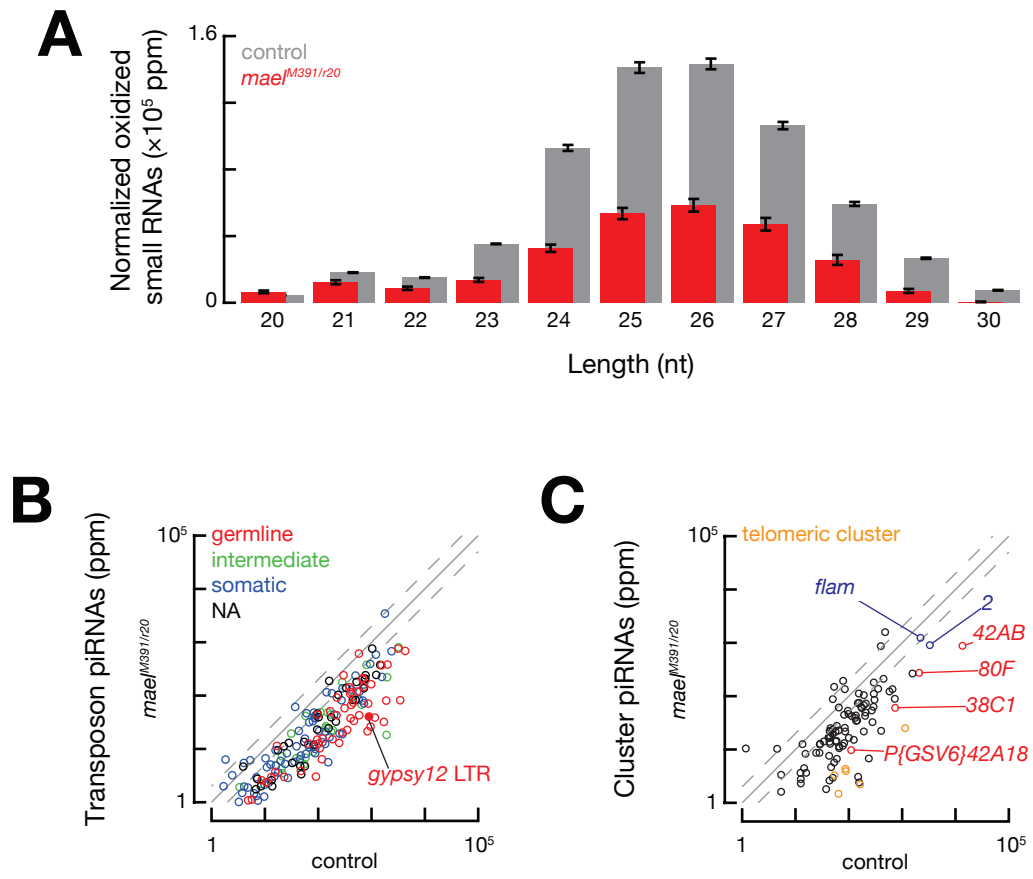

A
Normalized oxidized small RNAs (×10^5 ppm)
1.6
0
control
mael^M391/r20
20
21
22
23
24
25
26
27
28
29
30
Length (nt)
B
Transposon piRNAs (ppm)
mael^M391/r20
10^5
1
germline
intermediate
somatic
NA
gypsy12 LTR
1
10^5
control
C
Cluster piRNAs (ppm)
mael^M391/r20
10^5
telomeric cluster
flam
2
42AB
80F
38C1
P{GSV6}42A18
1
1
10^5
control

Figure 2.10: Fewer piRNAs in *mael*$^{M391/r20}$ mutants

(A) A length distribution of normalized oxidized sRNA libraries (*n* = 3) expressed in wild-type control (gray) and *mael*$^{M391/r20}$ ovaries (red) that uniquely mapped to the genome. Standard deviations are displayed as error bars. (B) Scatter plot comparing unique piRNAs that mapped to germline (red), soma (blue), intermediate (green), or unknown (black) transposons. piRNAs mapping to *gypsy12* LTRs are labeled shown as a solid red circle. (C) Scatter plot comparing piRNAs that unambiguously mapped to piRNA clusters. Labeled clusters are uni- (blue) or dual-strand (red). Telomeric piRNA clusters are displayed as orange circles. The hashed grey line signifies a ≥2-fold change.

The decreased abundance of piRNAs from dual-strand clusters in *mael*$^{M391/r20}$ mutants was accompanied by a marked loss of piRNA amplification by the ping-pong cycle (Figures 2.11A and 2.11B). Despite the loss of piRNAs produced by ping-pong in ovaries without Mael, significant ping-pong was detected among total piRNAs (*mael*, Z_{10} = 59; control, Z_{10} = 67; Figure 2.11A). In *mael*$^{M391/r20}$ ovaries, however, significant ping-pong was not detected among piRNAs mapping unambiguously to clusters (*mael*, Z_{10} = 0.58; control, Z_{10} = 23; excluding uni-strand clusters), whereas piRNAs mapping outside of clusters continued to be amplified (*mael*, Z_{10} = 8.1; control, Z_{10} = 19; Figure 2.11B). In *mael*$^{M391/r20}$ ovaries, outside of clusters, ping-pong can be attributed to piRNAs that uniquely mapped to *R1* LTR retrotransposons, which produced >2-fold more piRNAs than control ovaries.

Figure 2.11

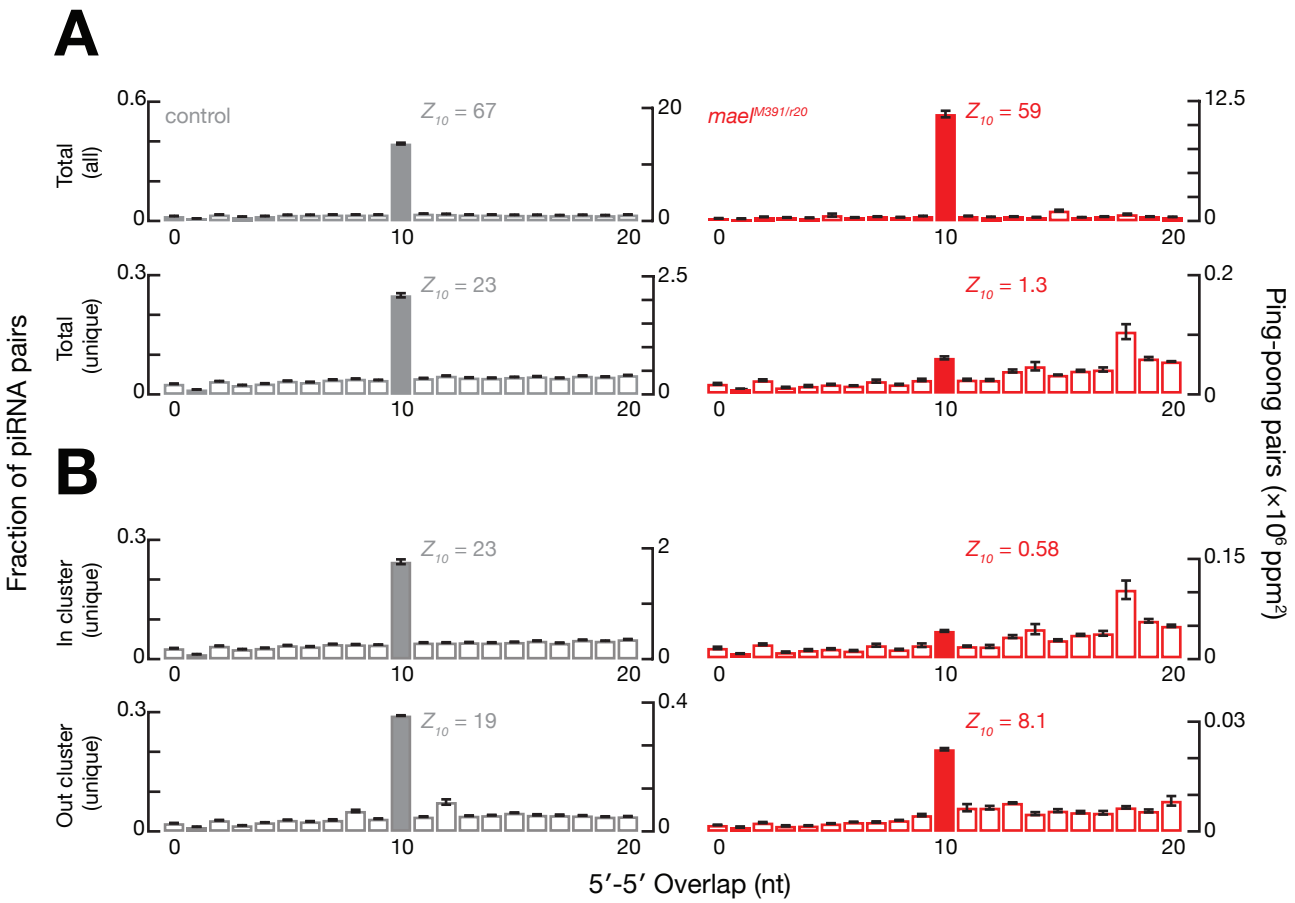

A
B
control
mael M391/r20
Z10 = 67
Z10 = 59
Z10 = 23
Z10 = 1.3
Z10 = 23
Z10 = 0.58
Z10 = 19
Z10 = 8.1
Total (all)
Total (unique)
In cluster (unique)
Out cluster (unique)
Fraction of piRNA pairs
Ping-pong pairs (×10⁶ ppm²)
5′-5′ Overlap (nt)

Figure 2.11: Fewer piRNAs produced by ping-pong without Mael

(A) Ping-pong analysis for all or uniquely mapping total piRNAs and (B) piRNAs that uniquely mapped in or out of dual-strand piRNA clusters. The left y-axis displays the fraction of piRNAs that overlapped on opposite genomic strands at each 5′ to 5′ distance. The right y-axis displays the number of ping-pong pairs detected. The value for the 10-nt overlap is shown as a solid bar. All data are the mean of three biological replicates from oxidized piRNAs from *mael*$^{M391/r20}$ (red) or wild-type control (gray) ovaries. Standard deviations are displayed as error bars.

While ping-pong remained significant in *mael*$^{M391/r20}$ ovaries fewer piRNAs were produced by ping-pong compared to wild-type control. This suggests that either the ping-pong machinery is normal and that piRNA precursors are not available for processing or that ping-pong has become inefficient. Because *mael*$^{M391/r20}$ ovaries appear to have abundant potential piRNA precursors (Figures 2.5 and 2.6), we looked at the expression of piRNA pathway genes that most affect ping-pong (Han et al., 2015). Although several genes had small but significant changes in transcript abundance in *mael*$^{M391/r20}$ mutant ovaries, *ago3* transcripts decreased 11 ± 2-fold (p = 0.001; Table 2.1). The decrease in *ago3* was not unique to *mael*$^{M391/r20}$ mutants. *ago3* transcripts decreased 7 ± 2-fold (p = 0.001) in *piwi*$^{2/Nt}$ mutants, 2 ± 1-fold (p = 0.006) in *armi*$^{72.1/G728E}$ mutants, and 4 ± 1-fold (p = 0.003) in *rhi*$^{2/KG}$ mutants. Because *ago3* is a heterochromatic gene with large transposon-filled introns, it is possible that *ago3* transcription could be disrupted when transposons are derepressed. Consistent with this hypothesis, several transposons in *ago3* introns in *mael*$^{M391/r20}$ ovaries had increased transcripts (Figure 2.12A), nascent transcription (Figure 2.12B), and more H3K4me3 signal (Figure 2.12C).

Table 2.1

	***mael*$^{M391/r20}$**		***piwi*$^{2/Nt}$**		***armi*$^{72.1/G728E}$**		***rhi*$^{2/KG}$**	
gene	**FC (*mael*/control)**	***p*-value**	**FC**	***p*-value**	**FC**	***p*-value**	**FC**	***p*-value**
ago3	0.09 ± 0.02	1.2×10^{-3}	0.14 ± 0.03	1.4×10^{-3}	0.41 ± 0.08	5.9×10^{-3}	0.27 ± 0.08	3.1×10^{-3}
aub	1.12 ± 0.08	0.038	1.10 ± 0.10	0.093	1.40 ± 0.10	7.5×10^{-4}	1.00 ± 0.10	0.72
cuff	1.56 ± 0.09	1.2×10^{-4}	1.80 ± 0.20	2.0×10^{-3}	1.90 ± 0.20	5.5×10^{-4}	1.40 ± 0.30	0.062
krimp	0.60 ± 0.10	0.022	0.70 ± 0.10	0.021	0.70 ± 0.10	0.025	0.80 ± 0.20	0.18
qin	1.10 ± 0.20	0.25	1.20 ± 0.20	0.070	1.60 ± 0.20	2.6×10^{-3}	0.90 ± 0.10	0.37
rhi	0.60 ± 0.10	0.014	0.80 ± 0.10	0.055	0.80 ± 0.10	0.11	0.10 ± 0.02	1.0×10^{-3}
spn-E	1.00 ± 0.20	0.69	1.10 ± 0.20	0.51	1.10 ± 0.20	0.28	1.30 ± 0.20	0.082
vas	1.10 ± 0.30	0.75	1.40 ± 0.20	0.018	1.40 ± 0.30	0.022	1.90 ± 0.30	1.1×10^{-3}
vret	0.90 ± 0.10	0.087	0.85 ± 0.08	0.046	0.89 ± 0.08	0.1	0.90 ± 0.10	0.15

Table 2.1: Expression of ping-pong related genes in *mael*, *piwi*, *armi*, and *rhi* mutants

Table displaying the fold change (mutant/control) of RNA-seq reads mapping to genes involved with ping-pong (Han et al., 2015). *p*-values were calculated from three biological replicates using an unpaired, two-tailed t-test.

Figure 2.12

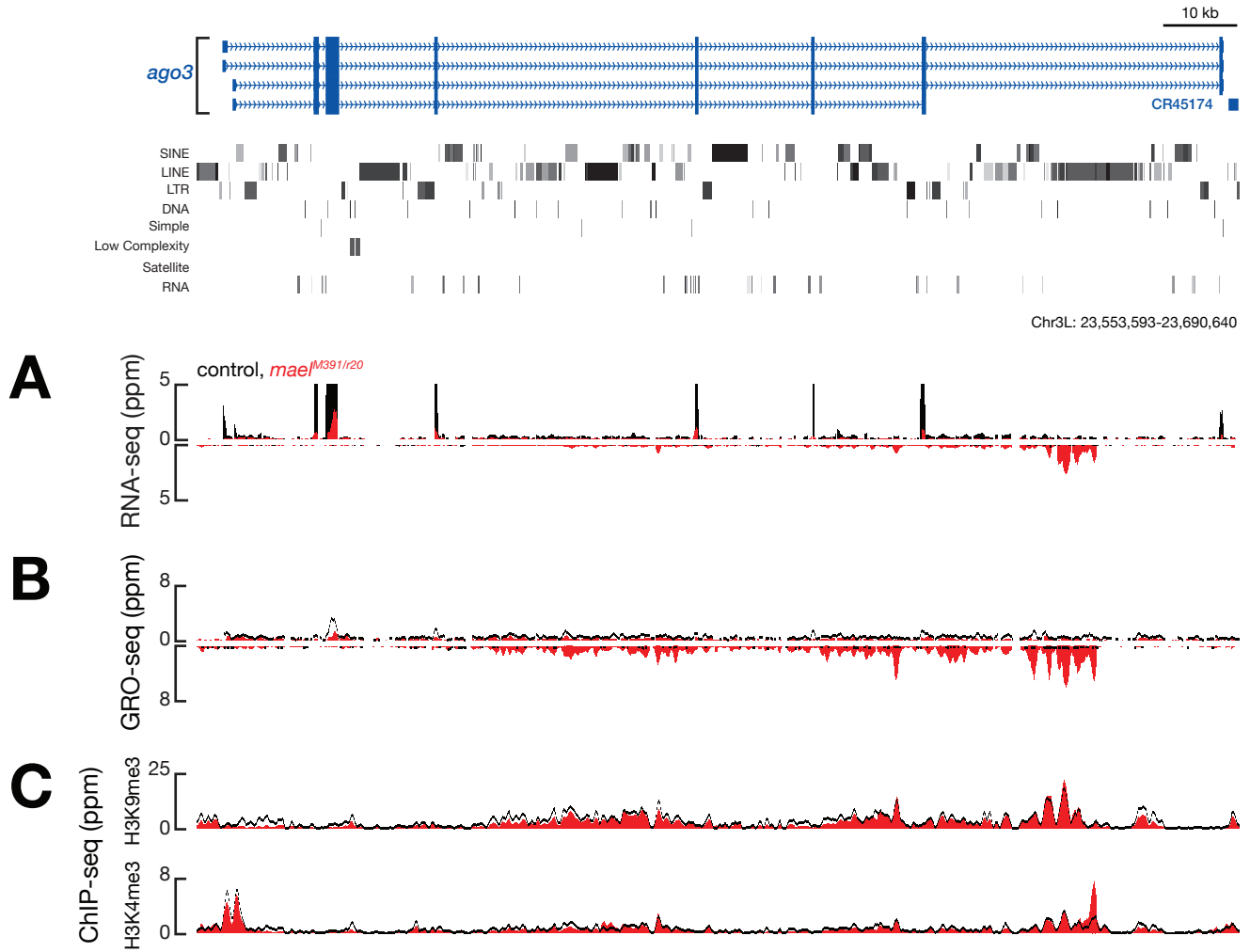

10 kb
ago3
CR45174
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Chr3L: 23,553,593-23,690,640
A
control, mael M391/r20
RNA-seq (ppm)
5
0
5
B
GRO-seq (ppm)
8
0
8
C
ChIP-seq (ppm)
H3K9me3
25
0
H3K4me3
8
0

Figure 2.12: Transposons in *ago3* introns are active in *mael*$^{M391/r20}$ ovaries

(A) RNA-seq, (B) GRO-seq, and (C) H3K9me3 and H3K4me3 ChIP-seq profiles for the *ago3* genomic region from *mael*$^{M391/r20}$ (red) and control (black) ovaries. Genes and their isoforms are displayed in blue while RepeatMasker sequences are displayed below.

Like piRNA production by the ping-pong cycle, the machinery required to produce phased piRNAs appears unaltered in *mael*$^{M391/r20}$ mutants. Despite the reduced abundance of total, uniquely mapping, phased piRNAs in *mael*$^{M391/r20}$ ovaries (Figure 2.13A), significant tail-to-head piRNA phasing remains (total piRNAs: *mael*, Z_1 = 11 versus control, Z_1 = 14; Figure 2.13B). While ping-pong was not detected among cluster-mapping piRNAs, phasing remained significant among piRNAs that uniquely mapped both inside and outside piRNA clusters (in cluster: *mael*, Z_1 = 9.0 versus control, Z_1 = 12; out cluster: *mael*, Z_1 = 14 versus control, Z_1 = 16; Figure 2.13B). We conclude that the primary piRNA biogenesis machinery is functional in the absence of Mael.

Figure 2.13

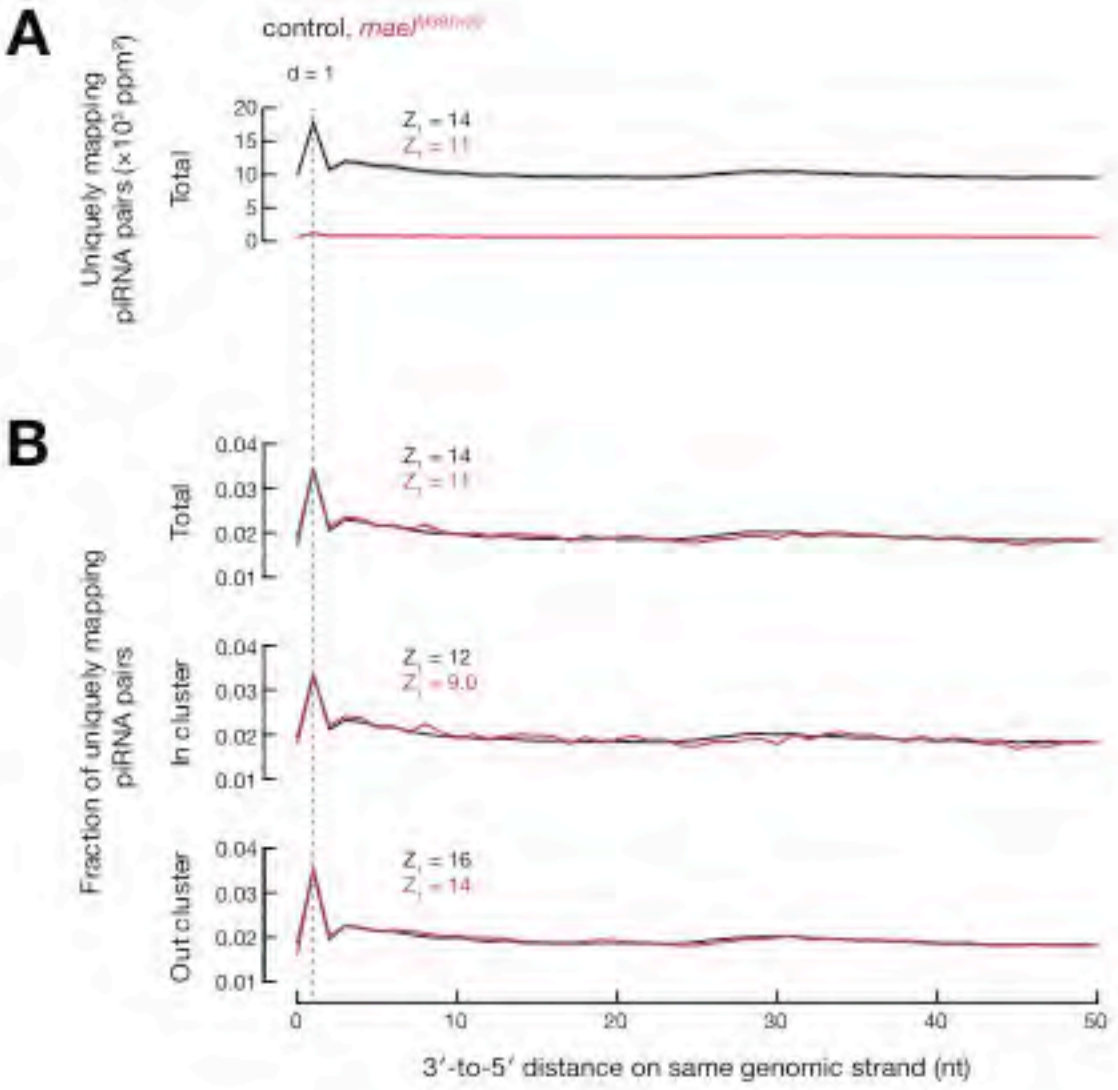

A
control, mael
d = 1
Uniquely mapping piRNA pairs (×10³ ppm²)
Total
20
15
10
5
0
Z = 14
Z = 11
B
Fraction of uniquely mapping piRNA pairs
Total
0.04
0.03
0.02
0.01
Z = 14
Z = 11
In cluster
Z = 12
Z = 9.0
Out cluster
Z = 16
Z = 14
0
10
20
30
40
50
3′-to-5′ distance on same genomic strand (nt)

Figure 2.13: Phasing is functional without Mael

(A) Total number of piRNA pairs that are found at each distance from the 3′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs on the same genomic strand. (B) Frequency of distances from the 3′ ends to the 5′ ends of piRNAs found in or out of piRNA clusters on the same genomic strand. All data are the mean of three biological replicates from uniquely mapping oxidized piRNAs from *mael*$^{M391/r20}$ (red) or wild-type control (black) ovaries.

Armi and Piwi also Repress Transcription in Dual-Strand Clusters

The current model for piRNA biogenesis in flies places Armi upstream and Mael downstream of Piwi (Malone et al., 2009; Haase et al., 2010; Saito et al., 2010; Sienski et al., 2012; Czech et al., 2013; Pandey et al., 2017; Rogers et al., 2017). Throughout this study, we used the transheterozygous *piwi* mutant, $piwi^{2/Nt}$, which has one copy of Piwi that can still be loaded with a piRNA but cannot enter the nucleus for transcription repression (Klenov et al., 2011). Consistent with this model, loss of either Armi or nuclear Piwi phenocopies the loss of Mael. For example, cluster-mapping transcripts expression increased to similar levels in $armi^{72.1/G728E}$, $piwi^{2/Nt}$, $mael^{M391/r20}$ ovaries compared to control (Figure 2.14A). Like $mael^{M391/r20}$ mutants, *gypsy12* LTRs increase >13-fold ($p = 2.5 \times 10^{-4}$) and >9-fold ($p = 9.3 \times 10^{-5}$) in $armi^{72.1/G728E}$ and $piwi^{2/Nt}$ ovaries, respectively (Figure 2.15A). Furthermore, in $armi^{72.1/G728E}$ and $piwi^{2/Nt}$ ovaries, *cluster38C1* transcripts were >20-fold ($p = 1.6 \times 10^{-4}$) and >13-fold ($p = 6.4 \times 10^{-6}$) more abundant than control ovaries, respectively (Figure 2.15B). While Armi was required for *cluster38C1* piRNAs (110 ± 60-fold decrease compared to control, $p = 0.002$), $piwi^{2/Nt}$ ovaries had normal levels of *cluster38C1* piRNAs ($p = 0.53$; Figure 2.14B). This is consistent with the possibility that the increase in *cluster38C1* transcripts in $piwi^{2/Nt}$ ovaries was not due to the loss of downstream piRNA processing and that Piwi may also be required to repress transcription of *cluster38C1*.

Figure 2.14

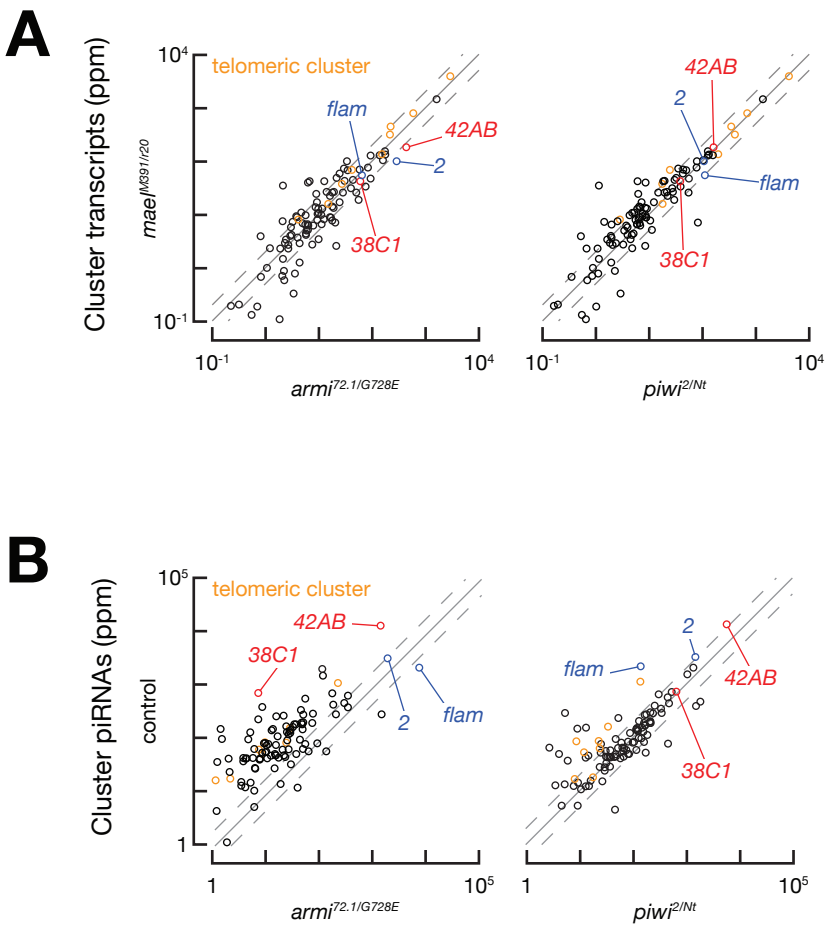

A
Cluster transcripts (ppm)
mael^IM891/r20
telomeric cluster
flam
42AB
2
38C1
armi^72.1/G728E
piwi^2/Nt
B
Cluster piRNAs (ppm)
control
telomeric cluster
42AB
38C1
2
flam
armi^72.1/G728E
piwi^2/Nt

Figure 2.14: More piRNA cluster transcripts in *armi*$^{72.1/G728E}$ and *piwi*$^{2/Nt}$ mutants than control ovaries

(A) Scatter plot comparing transcripts and (B) piRNAs that uniquely mapped to piRNA clusters from *mael*$^{M391/r20}$ or control to *armi*$^{72.1/G728E}$ or *piwi*$^{2/Nt}$ mutant ovaries. Labeled clusters are uni- (blue) or dual-strand (red). Telomeric piRNA clusters are displayed as orange circles. The hashed grey line signifies a ≥2-fold change.

Figure 2.15

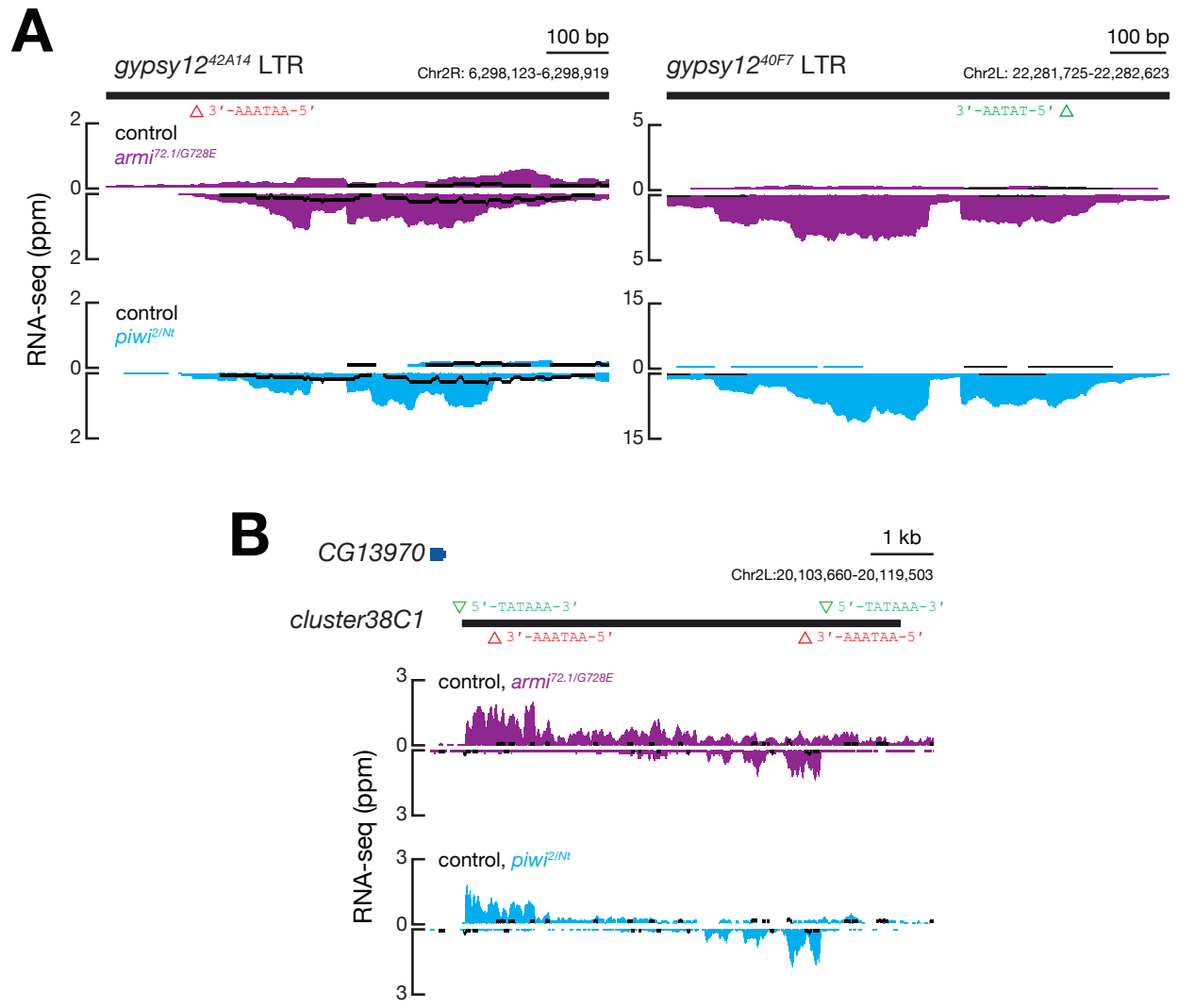
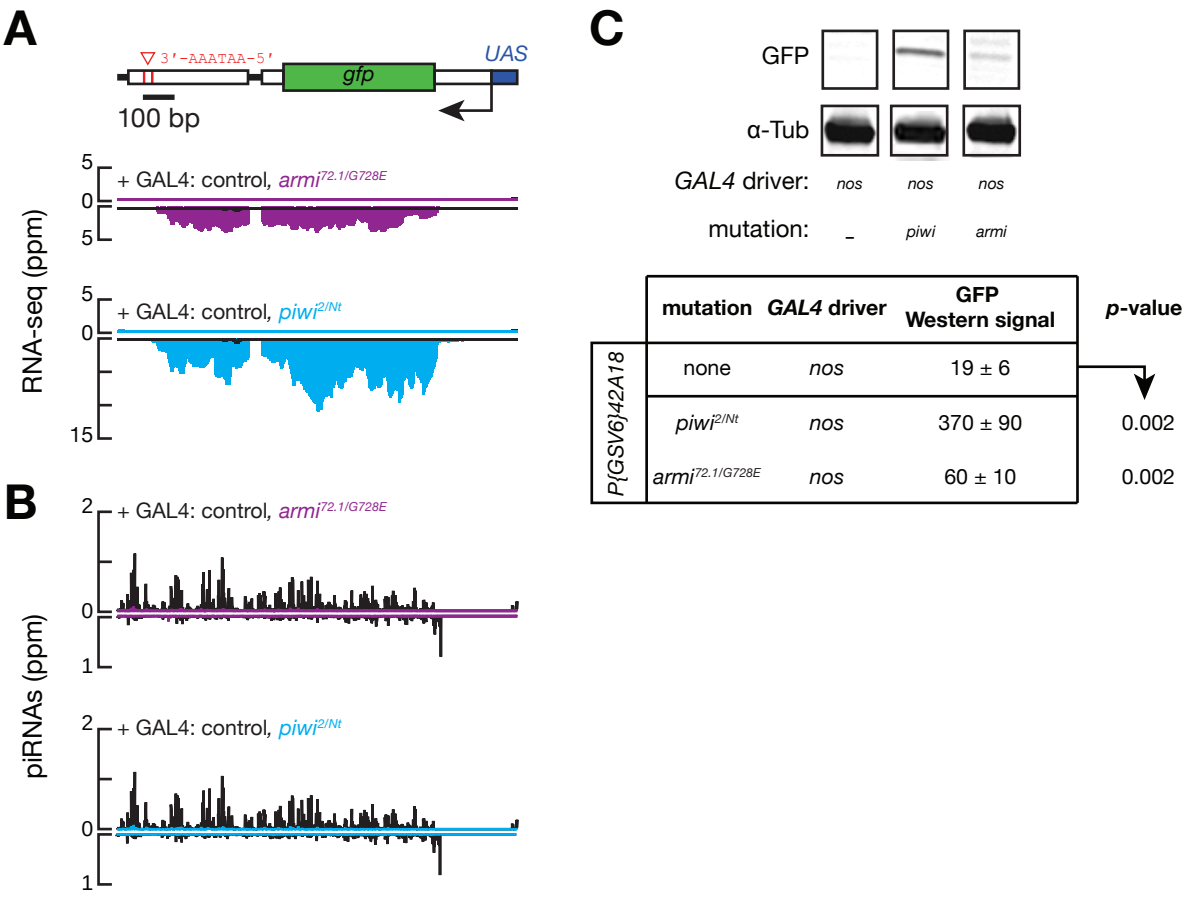

A
100 bp
gypsy12^42A14 LTR
Chr2R: 6,298,123-6,298,919
3′-AAATAA-5′
control
armi^72.1/G728E
control
piwi^2/Nt
RNA-seq (ppm)
100 bp
gypsy12^40F7 LTR
Chr2L: 22,281,725-22,282,623
3′-AATAT-5′
B
CG13970
1 kb
Chr2L:20,103,660-20,119,503
cluster38C1
5′-TATAAA-3′
3′-AAATAA-5′
control, armi^72.1/G728E
control, piwi^2/Nt
RNA-seq (ppm)

Figure 2.15: *gypsy12* LTRs and *cluster38C1* are derepressed in *armi*$^{72.1/G728E}$ and *piwi*$^{2/Nt}$ ovaries

(A) RNA-seq profiles for *gypsy12*^{42A14} (left), *gypsy12*^{40F7} (right), and (B) *cluster38C1* from *armi*$^{72.1/G728E}$ (purple), *piwi*$^{2/Nt}$ (cyan), and control (black) ovaries. TSSs and TTSs are marked with green and red triangles, respectively.

Like *cluster38C1*, *P{GSV6}42A18* also uses canonical Pol II promoters and loss of either Armi or nuclear Piwi allowed transcription of a spliced *gfp* mRNA that could be translated (Figures 2.16A and 2.16C). Finally, consistent with our previous results, $armi^{72.1/G728E}$, $piwi^{2/Nt}$, and $mael^{M391/r20}$ mutant ovaries all had fewer piRNAs than control (Figure 2.16B). Our results are consistent with Armi, Piwi, and Mael all functioning in the same pathway to repress canonical transcription in dual-strand piRNA clusters.

Figure 2.16



	mutation	*GAL4* driver	GFP Western signal	*p*-value
P{GSV6}42A18	none	*nos*	19 ± 6	
	piwi$^{2/Nt}$	*nos*	370 ± 90	0.002
	armi$^{72.1/G728E}$	*nos*	60 ± 10	0.002

Figure 2.16: *P{GSV6}42A18* is active in *armi$^{72.1/G728E}$* and *piwi$^{2/Nt}$* ovaries

(A) RNA-seq and (B) piRNA profiles for *P{GSV6}42A18* from *armi$^{72.1/G728E}$* (purple) and *piwi$^{2/Nt}$* (cyan) mutant ovaries also expressing germline GAL4-VP16. (C) Western blots and quantification for GFP normalized to α-Tub from *armi$^{72.1/G728E}$*, *piwi$^{2/Nt}$*, and control ovaries. *p*-values were measured using an unpaired, two-tailed t-test compared to *w^{1118}; P{GSV6}42A18*/*nos-gal4*; +.

Rhi is Not Required to Repress Transcription in Dual-Strand Clusters

While Armi, Piwi, and Mael are required to repress canonical Pol II transcription in dual-strand clusters, Rhi is required for non-canonical transcription of dual-strand clusters (Klattenhoff et al., 2009; Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014; Andersen et al., 2017). However, without Rhi, it is possible that dual-strand cluster transcription may be canonically initiated and terminated. Consistent with this possibility, in *rhi*$^{2/KG}$ ovaries, *gypsy12* LTRs increase >45-fold ($p = 2.6 \times 10^{-4}$) yet transcription terminates ~400 bps downstream with or without an AATAAA cleavage and poly(A) consensus sequence, similar to *mael*$^{M391/r20}$ mutants (Figure 2.17A). Moreover, like the two *gypsy12* LTRs, without Rhi, transcription of *cluster38C1* was able to initiate at the flanking promoters, but also terminated ~400 bps downstream near poly(A) cleavage sites (Figure 2.17B). While transcription can initiate without Rhi at promoters at *gypsy12*^{42A14}, *gypsy12*^{40F7}, and *cluster38C1*, we did not detect *gfp* transcripts or GFP protein in *rhi*$^{2/KG}$ ovaries (Figures 2.18A and 2.18B). Transcription was unable to continue past poly(A) cleavage sites in *rhi*$^{2/KG}$ mutant ovaries likely due to the loss of Cuff localization (Mohn et al., 2014; Chen et al., 2016). Consistent with this possibility, germline depletion of Cuff phenocopied *rhi*$^{2/KG}$ mutants (Figures 2.17A, 2.17B, 2.18A, and 2.18B). Thus, the *UAS* enhancer appeared to remain effectively silenced by heterochromatin without Rhi. In summary, both the flanking *cluster38C1* promoters and the *gyspy12* LTR can function in heterochromatin without Rhi. In wildtype germline nurse cells, however, Rhi ensures that *gypsy12*^{42A14}, *gypsy12*^{40F7}, and *cluster38C1* are transcribed non-canonically via Moon and Cuff subsequently processed into piRNAs. We conclude that Rhi is not required for repression of canonical transcription dual-strand piRNA clusters.

Figure 2.17

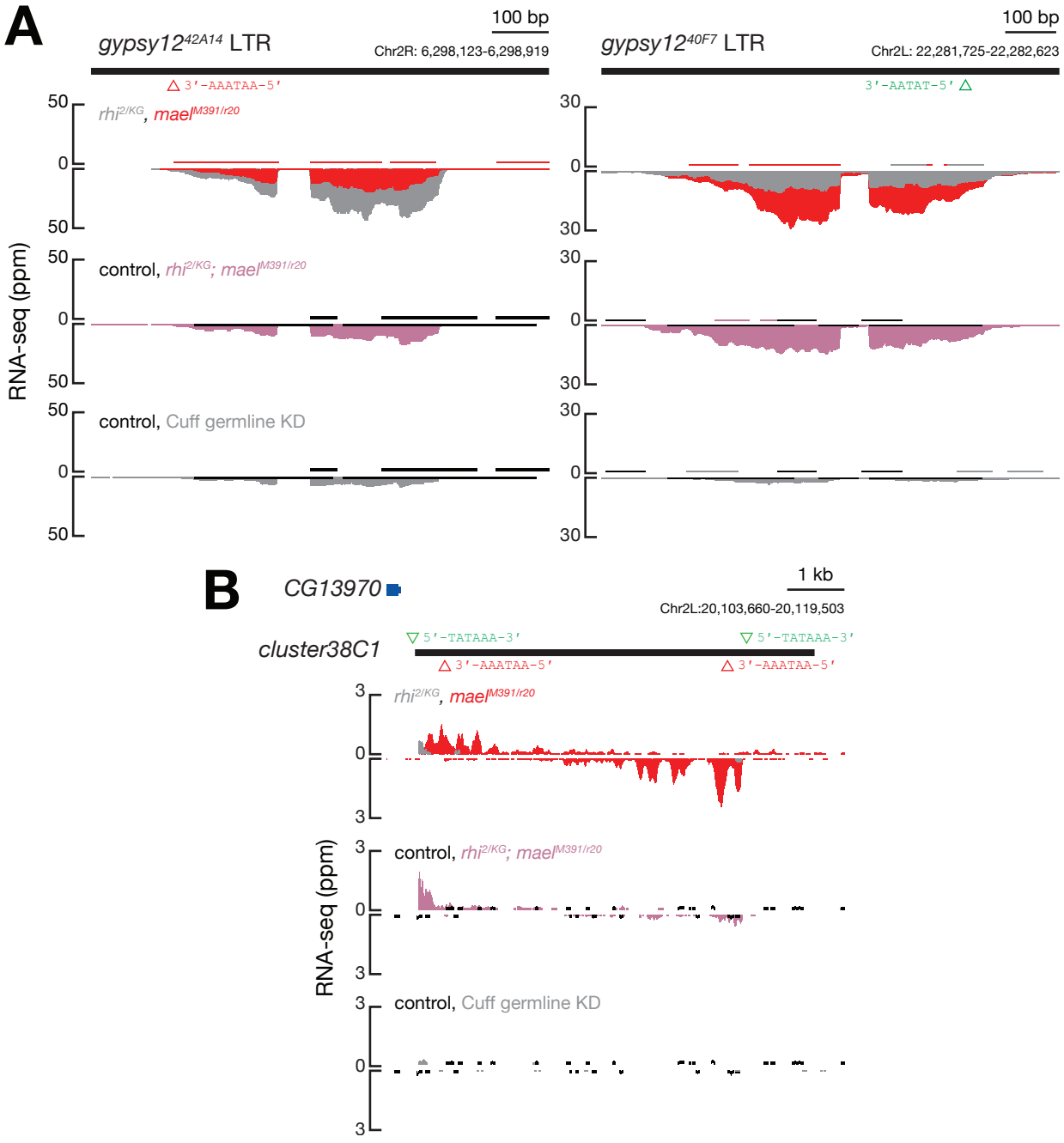
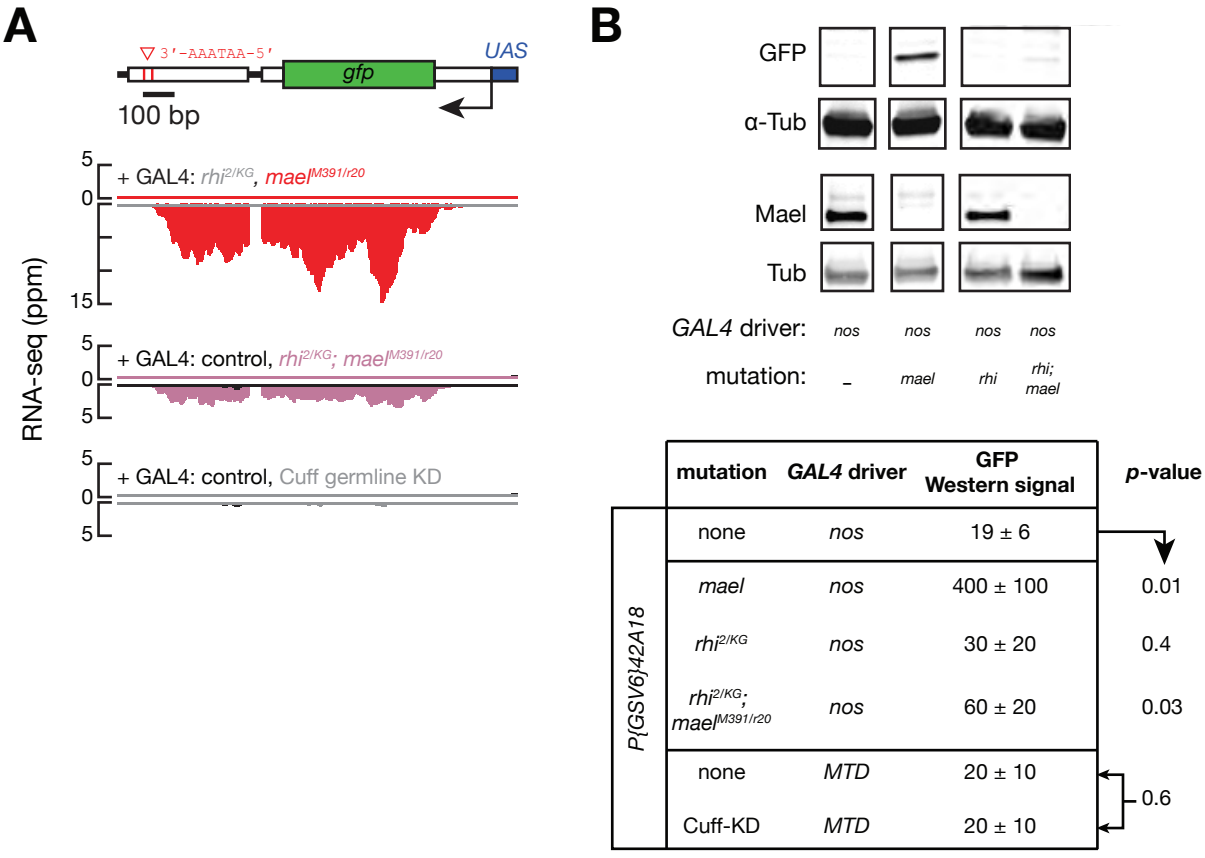

A
gypsy12^42A14 LTR
100 bp
Chr2R: 6,298,123-6,298,919
3′-AAATAA-5′
gypsy12^40F7 LTR
100 bp
Chr2L: 22,281,725-22,282,623
3′-AATAT-5′
rhi^2/KG, mael^M391/r20
control, rhi^2/KG; mael^M391/r20
control, Cuff germline KD
RNA-seq (ppm)
B
CG13970
1 kb
Chr2L:20,103,660-20,119,503
cluster38C1
5′-TATAAA-3′
3′-AAATAA-5′
rhi^2/KG, mael^M391/r20
control, rhi^2/KG; mael^M391/r20
control, Cuff germline KD
RNA-seq (ppm)

Figure 2.17: Transcription can initiate at *gypsy12* and *cluster38C1* without Rhi

(A) RNA-seq profiles for *gypsy12*^{42A14} (left) *gypsy12*^{40F7} (right) and (B) *cluster38C1* from *rhi*$^{2/KG}$, *mael*$^{M391/r20}$, *rhi*$^{2/KG}$*; mael*$^{M391/r20}$ double mutant, Cuff germline knockdown, and control ovaries. TSSs and TTSs are marked with green and red triangles, respectively.

Figure 2.18



mutation	*GAL4* driver	GFP Western signal	*p*-value
none	*nos*	19 ± 6	
mael	*nos*	400 ± 100	0.01
$rhi^{2/KG}$	*nos*	30 ± 20	0.4
$rhi^{2/KG}$; $mael^{M391/r20}$	*nos*	60 ± 20	0.03
none	*MTD*	20 ± 10	0.6
Cuff-KD	*MTD*	20 ± 10	

P{GSV6}42A18

Figure 2.18: Rhi is not required to repress transcription of *P{GSV6}42A18*

A) RNA-seq profiles for *P{GSV6}42A18* from $rhi^{2/KG}$, $mael^{M391/r20}$, $rhi^{2/KG}$; $mael^{M391/r20}$ double mutant, Cuff germline knockdown, and control ovaries also expressing germline GAL4-VP16. (B) Western blots and quantification for GFP normalized to α-Tub. *p*-values for genetic mutants were measured using an unpaired, two-tailed t-test compared to w^{1118}; *P{GSV6}42A18*/*nos-gal4*; + or to MTD-GAL4 control for Cuff germline knockdown.

One explanation for the absence of canonical transcription of *gfp* in *P{GSV6}42A18* in *rhi*$^{2/KG}$ ovaries is that Rhi is required to create a transcriptionally permissive environment in dual-strand piRNA clusters. To test this hypothesis, we used RNAi to reduce, but not eliminate Rhi in the germline (Figure 2.19A). Compared to wild-type control ovaries, both *rhi*$^{2/KG}$ mutants and Rhi germline knockdown had a ~10-fold loss of *rhi* transcript. In Rhi-depleted ovaries, however, low levels of RNA-seq reads were detected across the entire *rhi* coding sequence whereas in *rhi*$^{2/KG}$ mutants, RNA-seq reads sharply fell ~200 bp after the *rhi* TSS. Furthermore, while germline depletion of Rhi led to the loss of nuclear foci, diffuse nuclear Rhi was still detected (Figure 2.19B). In contrast, *rhi*$^{2/KG}$ ovaries only had background levels of nuclear signal. Consistent with this hypothesis, we detected spliced *gfp* mRNA in Rhi germline knockdown flies, unlike *rhi*$^{2/KG}$ genetic null mutants (Figure 2.19C).

Figure 2.19

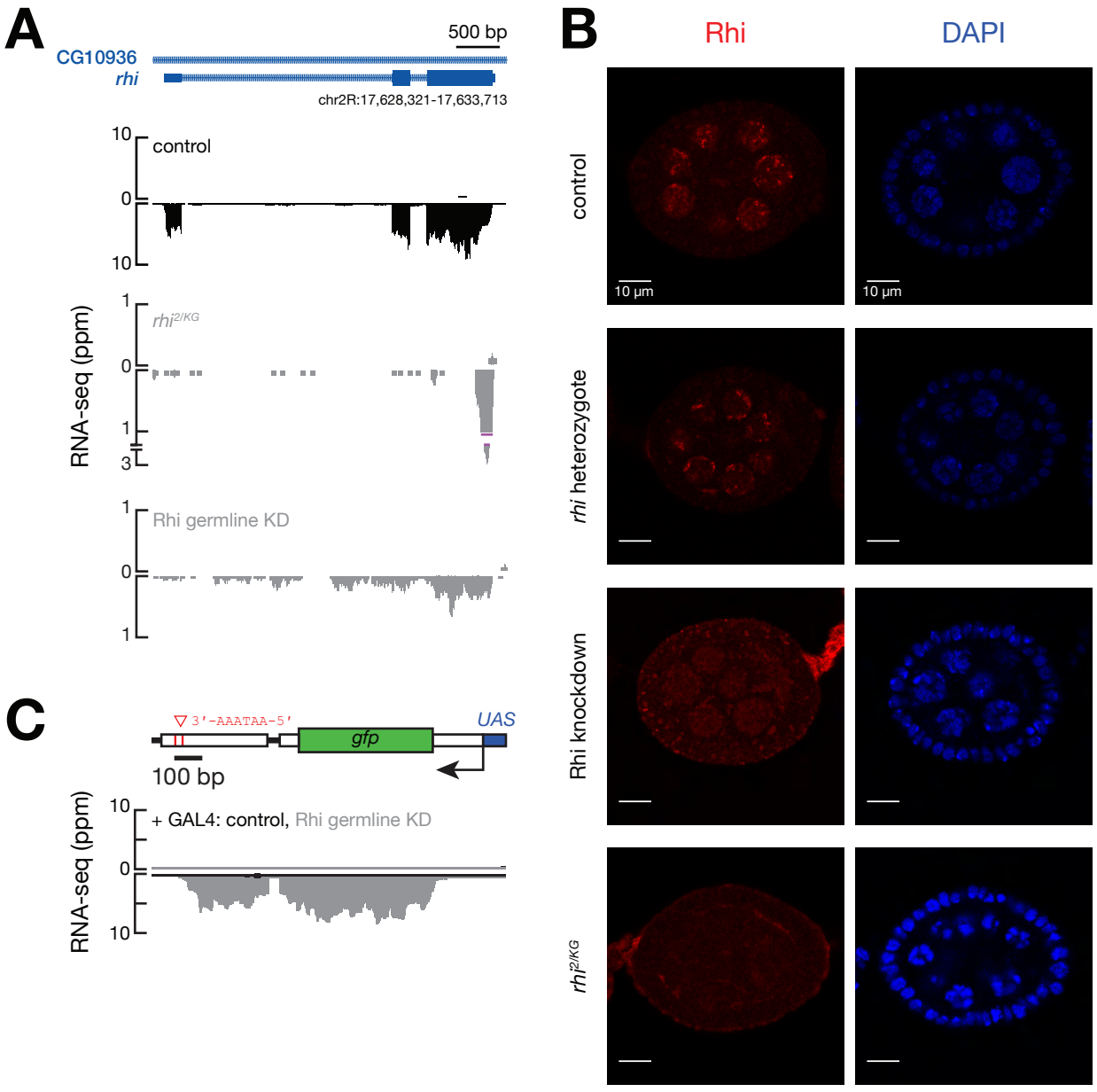

A
500 bp
CG10936
rhi
chr2R:17,628,321-17,633,713
control
rhi2/KG
Rhi germline KD
RNA-seq (ppm)
B
Rhi
DAPI
control
rhi heterozygote
Rhi knockdown
rhi2/KG
10 µm
C
3'-AAATAA-5'
UAS
gfp
100 bp
+ GAL4: control, Rhi germline KD
RNA-seq (ppm)

Figure 2.19: Rhi promotes expression of *P{GSV6}42A18*

(A) RNA-seq profiles for the *rhi* genomic region from control, $rhi^{2/KG}$ mutant, and Rhi germline knockdown ovaries. Genes are displayed above in blue. (B) Immuno-detection of Rhi protein (left, red) in wild-type, *rhi* heterozygote, $rhi^{2/KG}$ mutant, and Rhi-depleted ~stage 4 egg chambers. Gray scale bars are 10 μm. DAPI, 4′,6-diamidino-2-phenyl-indole (left, blue). (C) RNA-seq profiles for *P{GSV6}42A18* from Rhi-depleted and control ovaries expressing germline GAL4-VP16.

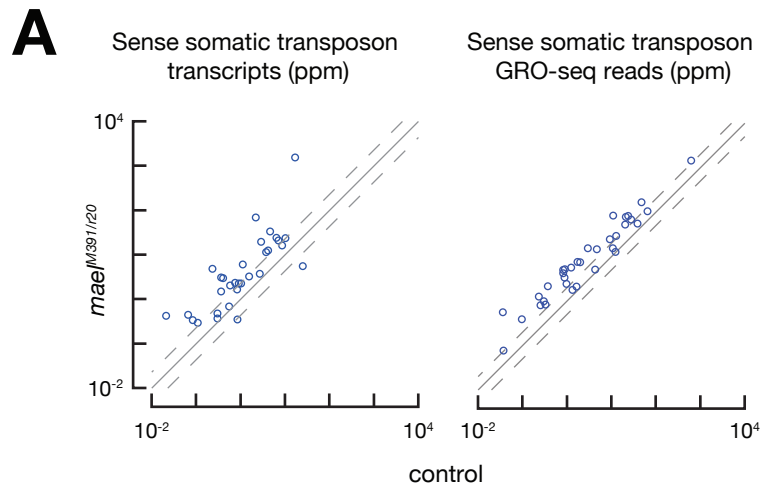
Because Rhi promotes transcription while Mael represses transcription, loss of Rhi should rescue transcription repression in dual-strand piRNA clusters in *mael*$^{M391/r20}$ mutants. Expectedly, because the *gypsy12* LTRs in *42AB* and *cluster62* are active in both *rhi*$^{2/KG}$ and *mael*$^{M391/r20}$ ovaries, *gypsy12*^{42A14} and *gypsy12*^{40F7} expression did not change in *rhi*$^{2/KG}$*; mael*$^{M391/r20}$ double mutants (Figure 2.17A). In *cluster38C1*, however, further loss of Rhi in *mael*$^{M391/r20}$ ovaries led to a >3-fold decrease in *cluster38C1* transcripts compared to *mael*$^{M391/r20}$ single mutants (Figure 2.17B). In addition, similar to *rhi*$^{2/KG}$ single mutants, *cluster38C1* transcripts in *rhi*$^{2/KG}$*; mael*$^{M391/r20}$ double mutants also terminated at poly(A) cleavage sites. Finally, *rhi*$^{2/KG}$*; mael*$^{M391/r20}$ ovaries produced 3 ± 1-fold fewer (p = 0.044) *gfp* transcripts and 6 ± 3-fold fewer (p = 0.021) GFP protein than *mael*$^{M391/r20}$ single mutants (Figures 2.18A and 2.18B). Our data suggest that independent of Rhi, canonical Pol II transcription is repressed by Mael. Altogether, Rhi and Mael function in oppositional roles: Rhi promotes transcription–both canonical and non-canonical–while Mael, guided by Piwi, represses canonical transcription in dual-strand piRNA clusters.

Mael Represses Transcription of Heterochromatin in the Ovary Somatic Follicle Cells

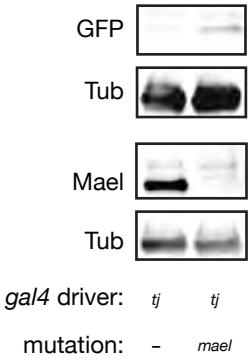
Outside of germ cells, Mael is also required to repress transposons in the somatic follicle cells, which support embryo development (Sienski et al., 2012; Muerdter et al., 2013; Figure 2.20A). Similar to germline piRNA clusters, *mael*$^{M391/r20}$ ovaries also had fewer piRNAs mapping to somatic piRNA-producing loci *flam* (*mael*/control = 0.33 ± 0.03; $p = 1.8 \times 10^{-5}$) and *traffic jam* (*mael*/control = 0.53 ± 0.08; p = 0.0010; Figures 2.10A and 2.10B). Consistent with a somatic role for Mael, infertility in *mael*$^{M391/r20}$ mutants was rescued only when a FLAG-Mael transgene was simultaneously expressed in both the germline and somatic follicle cells (Table 2.2).

The piRNA pathway has long been associated with repressing genes and transposons using heterochromatin (Pal-Bhadra et al., 2004; Moshkovich and Lei, 2010; Todeschini et al., 2010; Sienski et al., 2012; Donertas et al., 2013; Gu and Elgin, 2013; Le Thomas et al., 2013; Le Thomas et al., 2014; Sienski et al., 2015; Yu et al., 2015b). Mael has also been linked to heterochromatin and silencing in the soma: Mael-depleted OSCs show minor changes in H3K9me3 but cannot repress somatic transposons (Sienski et al., 2012). Using the *traffic jam-gal4* driver to express *P{GSV6}42A18* in somatic follicle cells, more GFP was detected in $mael^{M391/r20}$ mutants (*mael*/control = 3.3 ± 0.9, $p = 0.0043$; Figure 2.20B). Therefore, as in the germline, in somatic follicle cells, Mael repressed transcription of a euchromatic transgene inserted into *42AB*.

Figure 2.20



B



	mutation	GAL4 driver	GFP Western signal	p-value
P{GSV6}42A18	none	*tj*	0.0046 ± 0.0008	4.3×10^{-3}
	mael$^{M391/r20}$	*tj*	0.015 ± 0.003	

Figure 2.20: Mael represses heterochromatic sequences in somatic follicle cells

(A) Scatter plots comparing sense RNA-seq (left) and GRO-seq (right) that unambiguously mapped to somatic transposons from *mael*$^{M391/r20}$ and control ovaries. (B) Western blots for GFP, Mael, and α-Tub from *mael*$^{M391/r20}$ and control ovaries expressing the somatic-follicle-cell-specific driver, *tj-gal4*. Quantification for GFP was normalized to α-Tub.

Table 2.2

Mutation	FLAG-Mael rescue	Dorsal appendages		Eggs per female per day	Hatch rate (%)
		N	Wild-type (%)		
mael$^{+/-}$	none	9976	96.6	50.18	53.8
mael$^{M391/r20}$	none	7	0.0	0.04	0
mael$^{M391/r20}$	*tj-gal4*	1394	2.0	3.63	0
mael$^{M391/r20}$	*nos-gal4*	1240	15.1	3.23	0.6
mael$^{M391/r20}$	*tj- + nos-gal4*	11,177	92.9	38.38	27.3
mael$^{M391/r20}$	*actin5C-gal4*	11,900	97.6	30.99	62.0

Table 2.2: Mael is required in both germline and somatic follicle cells for fertility

Table displaying the results of female fertility assays. Rescue using *nos* (germline), *tj* (somatic follicle cell), *tj* + *nos* (germline + somatic follicle cell), or *actin5C* (ubiquitous) drivers was assessed by wild-type dorsal appendage rate, eggs laid per female per day, and hatch rate. $mael^{+/-}$ are a mix of $mael^{M391/TM3,Sb}$ and $mael^{r20/TM3,Sb}$ females.

mael is a Suppressor of Position Effect Variegation

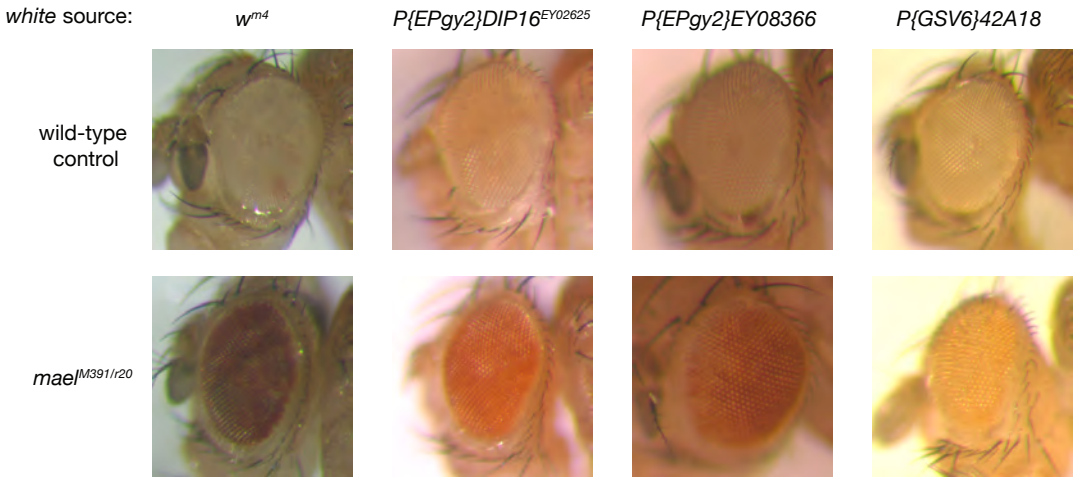
Position effect variegation (PEV) is the phenomenon in which gene expression is dependent on chromatin context. In the classic PEV example, w^{m4}, the euchromatic *white* gene, which is required for red pigment expression in the eyes, is silenced when it is moved near the centromere by a chromosomal inversion (Muller, 1930). Subsequent screens would reveal "suppressors of PEV," or mutations that would suppress the silencing of *white*. Suppressors of PEV were often heterochromatin related genes, such as *Suppressor of variegation 205* (*Su(var)2-5*), which encodes HP1a (Eissenberg et al., 1992). Thus, the loss of GFP silencing in $mael^{M391/r20}$ mutants is reminiscent of the mutation of a suppressor of PEV. In fact, *P{GSV6}42A18* contains a *mini-white* marker and was derepressed in $mael^{M391/r20}$ eyes compared to control as measured by red eye pigment (*mael*/control = 1.7 ± 0.4, $p = 0.02$; Figure 2.21).

Previously, the Lei lab used several *mini-white* transgene reporters inserted into heterochromatic regions near piRNA clusters as PEV reporters. While eye pigmentation was suppressed by mutating *Su(var)2-5* or *Su(var)3-9*, the reporter was not sensitive to *piwi* or *aub* mutations (Moshkovich and Lei, 2010). Using $P\{EPgy2\}DIP16^{EY02625}$, which is inserted upstream of *flam*, $mael^{M391/r20}$ flies also expressed more red pigment in the eyes (*mael*/control = 2.8 ± 0.7, $p = 2.6 \times 10^{-4}$; Figure 2.21). Furthermore, using *P{EPgy2}EY08366*, which is insertion near the proximal edge of *42AB*, loss of Mael caused a 9 ± 8-fold increase in red pigment expression ($p = 5.2 \times 10^{-4}$; Figure 2.21).

Finally, using the original w^{m4} mutation revealed that like *Su(var)2-5* and *Su(var)3-9*, loss of one copy of *mael* also suppressed the silencing of *white* ($mael^{M391}$/control = 100 ± 50, $p = 7.5 \times 10^{-6}$; $mael^{r20}$/control = 80 ± 40, $p = 1.4 \times 10^{-4}$; Table 2.3). Furthermore, in the eyes, Mael appears to repress *white* in a concentration-dependent manner: $mael^{M391/r20}$ mutants had ~2-fold more red pigment than $mael^{M391/+}$ or $mael^{r20/+}$ heterozygotes ($p = 1.1 \times 10^{-3}$ and 7.7×10^{-4}, respectively). Therefore, *mael*

is a suppressor of PEV in somatic tissue outside the ovaries and is required to repress the transcription of typically euchromatic genes silenced by heterochromatin.

Figure 2.21



white source	Mael?	OD480	p-value
w^{m4}	Yes	0.005 ± 0.002	2.9×10^{-6}
w^{m4}	No	0.70 ± 0.03	
P{EPgy2}DIP16^{EY02625}	Yes	0.007 ± 0.002	2.6×10^{-4}
P{EPgy2}DIP16^{EY02625}	No	0.018 ± 0.001	
P{EPgy2}EY08366	Yes	0.001 ± 0.001	5.2×10^{-4}
P{EPgy2}EY08366	No	0.011 ± 0.001	
P{GSV6}42A18	Yes	0.008 ± 0.001	1.6×10^{-3}
P{GSV6}42A18	No	0.014 ± 0.002	

Figure 2.21: *mael* is a suppressor of PEV

Representative images of fly eyes from *mael*$^{M391/r20}$ and wild-type control flies from different variegating strains. The source of *white* is shown above the image. In *w*^{m4}, *white* is near pericentromeric heterochromatin. *P{EPgy2}DIP16*^{EY02625}, *P{EPgy2}EY08366*, and *P{GSV6}42A18* all contain *mini-white*. *P{EPgy2}DIP16*^{EY02625} and *P{EPgy2}EY08366* are inserted near the *flam* and *42AB*, respectively. *P{GSV6}42A18* is inserted in *42AB*. Below the images are the quantification of red pigment from the different strains by measuring the optical density at 480 nm (OD480). *p*-values compare pigment from *mael*$^{M391/r20}$ and control eyes with the same *white* source.

Table 2.3

Genotype	OD480	*p*-value vs control
wm4; +; +	$5\times10^{-3} \pm 2\times10^{-3}$	control
wm4; +; mael$^{M391/+}$	0.49 ± 0.03	7.5×10^{-6}
wm4; +; mael$^{r20/+}$	0.39 ± 0.05	1.5×10^{-4}
wm4; +; mael$^{M391/r20}$	0.70 ± 0.03	2.9×10^{-6}
wm4; Su(var)2-5$^{5/+}$*; +*	0.60 ± 0.03	3.6×10^{-6}
wm4; +; Su(var)3-9$^{2/+}$	0.74 ± 0.02	2.2×10^{-7}

Table 2.3: *mael* is a strong suppressor of PEV

Table displaying the results of eye pigment assays measuring the OD480 from w^{m4} flies with different mutations. *p*-values for each genotype were calculated against the control, w^{m4}; +; +.

EXPERIMENTAL PROCEDURES

Dosophila Stocks

Mutant alleles, shRNA lines, GAL4-VP16 driver transgenes, and PEV lines have been described previously: *mael*$^{M391/r20}$ (Clegg et al., 1997), *rhi*^{02086} (*rhi*^{2}) and *rhi*^{KG00910} (*rhi*^{KG}) (Cogoni and Macino, 1999), *piwi*^{2} (Cox et al., 1998), *piwi*^{Nt} (Klenov et al., 2011), *armi*$^{72.1}$ (Cook et al., 2004), *arm*^{G728E} (Ozcan et al., 2015); shRNA lines targeting *rhi* (Bloomington #35171) and *cuff* (Bloomington #35182; Ni et al., 2011); *nanos*-GAL4 (Bloomington #4442), *tj*-GAL4 (DGRC #104055), and MTD-GAL4 (Bloomington #31777); *w*^{m4} (Bloomington #807; Muller, 1930), *P{EPgy2}DIP16*^{EY02625} (Bloomington #15577), and *P{EPgy2}EY08366* (Bloomington #19874). The *P{GSV6}42A18* transgene derives from *P{GSV6}*^{GS13456} (Chendrimada et al., 2005) and is located at Chr2R: 6,460,398-6,460,415. The FLAG-Mael rescue strain was a gift from Toshie Kai (Pek et al., 2009). Unless otherwise noted, all flies were in the *w*^{1118} background except *w*$^{+}$*; rhi*^{KG} strains. Both the *mael*^{M391} and *mael*^{r20} alleles were outcrossed for five generations.

General Methods

Stocks and crosses were grown at 25°C. Before dissection, flies were isolated 0–3 days after eclosion and given yeast paste for two days. Fly ovaries were then dissected and collected in 1× phosphate-buffered saline (1×PBS) [pH 7.4] (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) cooled on ice. Ovaries were then washed once with ice-cold 1×PBS. Total RNA was purified using mirVana (Life Technologies, #AM1561).

Unless otherwise stated, throughout this thesis, “wild-type control” samples matched the genotype of the mutant except for the mutant alleles. For example, the control for *w*^{1118}; *P{GSV6}42A18*/*nos-gal4*; *mael*$^{M391/r20}$ was *w*^{1118}; *P{GSV6}42A18*/*nos-gal4*; + not *w*^{1118}; *P{GSV6}42A18*/+; +.

Western Blotting

Ovary lysate was prepared essentially as described in (Tomari et al., 2004). After 1×PBS was removed, the ovaries were homogenized with a plastic pestle (Fisher Scientific, #12141364) in ice-cold lysis buffer (for each 100 mg ovaries, 100 µl of 100 mM potassium acetate, 30 mM HEPES-KOH [pH 7.4], 2 mM magnesium acetate, 1 mM dithiothreitol (DTT)) containing 1 mM AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride; EMD Millipore, #101500), 0.3 µM Aprotinin (Bio Basic Inc, #AD0153), 20 µM Bestatin (Sigma Aldrich,#B8385), 10 µM E-64 ((1S,2S)-2-(((S)-1-((4-Guanidinobutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamoyl)cyclopropanecarboxylic acid; VWR, #97063), and 10 µM Leupeptin (Fisher Scientific, #108975). Lysate was centrifuged at 13,000 × *g* for 30 min at 4°C and an equal volume of 2× loading dye (100 mM Tris-HCl [pH 6.8], 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, and 200 mM DTT) was added to the supernatant and heated to 95°C for 5 min.

The lysate was resolved through a 4–20% gradient polyacrylamide/SDS gel electrophoresis (Bio-Rad Laboratories, #5671085). After electrophoresis, proteins were transferred to a 0.45 µm pore polyvinylidene difluoride membrane (Millipore, #IPVH00010), the membrane blocked in Blocking Buffer (Rockland Immunochemicals, #MB-070) at room temperature for 1 h and then incubated overnight at 4°C in 1:1 Blocking Buffer:1×TBST [pH 7.5] (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.1% Tween 20 (v/v)) containing primary antibody (anti-GFP, Santa Cruz Biotechnology, #SC-9996, 1:2500 dilution; anti-α-Tubulin, DSHB, #12G10, 1:50,000 dilution, anti-Mael, gift from Julius Brennecke, 1:2500 dilution). The membrane was washed 3×5 min with 1×TBST [pH 7.5] at 25°C, incubated in 1:1 Blocking Buffer:1×TBST [pH 7.5] containing secondary antibody (donkey anti-mouse IRDye 680RD, LICOR Biosciences, #926-68072, 1:10,000 dilution; goat anti-mouse IRDye 800CW, LICOR Biosciences, #926-32210, 1:10,000 dilution) for 1 h at 25°C in the dark, and washed 5×10 min with

1×TBST [pH 7.5] at 25°C in the dark. The signal was detected using Odyssey Infrared Imaging System. Data were obtained from three independent biological replicates. Quantification was performed using Image Studio v4.0.21 (LI-COR). Mean values and standard deviation were compared between matching *mael*$^{M391/r20}$ and control flies. *p*-values were measured using an unpaired, two-tailed t-test.

General Overview of Computational Analyses

The analyses for sRNA-, RNA-, ChIP-, and GRO-seq libraries were performed using piPipes (v1.4; Han et al., 2014). All the libraries were mapped to the *D. Melanogaster* genome (dm6) obtained from the UCSC along with all the necessary annotations (http://hgdownload.cse.ucsc.edu/goldenPath/dm6/bigZips). Unique molecular identifiers (UMIs) were used as described in (Fu et al., 2018a). H3K4me3 ChIP-seq analysis used the psychENCODE pipeline (https://github.com/weng-lab/psychip_snakemake). Sequencing data are available from the NCBI BioProject Archive using accession number PRJNA448445.

Statistical Analysis

Statistical significance was computed using the R implementation (v3.4.0) of the student's t-test. The *p.adjust* function was used to correct the *p*-values for multiple testing using false discovery rate method.

Small RNA-seq Library Preparation and Analysis

Small RNA libraries were constructed as described {Han et al., 2015, #29200}. Briefly, total RNA (50 µg) was purified by 15% urea polyacrylamide gel electrophoresis (PAGE), selecting for 18–30 nt small RNAs. Half of the purified sRNAs were oxidized with $NaIO_4$ was used to deplete miRNAs and enrich for siRNAs and piRNAs (Li et al., 2009a). To reduce ligation bias, a 3′ adaptor with three random nucleotides at its 5′ end was used

(5′-rApp NNN TGG AAT TCT CGG GTG CCA AGG /ddC/-3′). 3′ adaptor was ligated using truncated, K227Q mutant T4 RNA Ligase 2 (made in lab) at 25°C for ≥16 h, sRNAs precipitated, and size selected as described in (Li et al., 2009a). To exclude 2S rRNA from sequencing libraries, 10 pmol 2S blocker oligo was added before 5′ adaptor ligation (Wickersheim and Blumenstiel, 2013). 5′ adaptor was added using T4 RNA ligase (Life Technologies, #AM2141) at 25°C for 2.5 h, followed by reverse-transcription using AMV reverse transcriptase (New England Biolabs, #M0277L) and PCR using AccuPrime *Pfx* DNA Polymerase (Invitrogen, #12344-024). Small RNA-seq libraries for three independent biological replicates were sequenced using a NextSeq500 (Illumina) to obtain 75 nt single-end reads.

Barcodes were sorted allowing one mismatch, and the 3′ adaptors, including the three random nucleotides, were identified and removed using the first ten nucleotides, allowing one mismatch. After adaptor removal, reads containing one or more nucleotides with Phred score <5 were discarded. sRNAs were first aligned to rRNA or miRNA hairpin sequences using Bowtie2 (v2.2.0; Langmead and Salzberg, 2012). Unaligned reads were mapped to the genome and 23–29 nt RNAs (fly piRNAs) were kept for analyses. The number of piRNAs overlapping each genomic feature (genes, transposons, and piRNA producing loci) were apportioned by the number of times they aligned to the genome.

Oxidized sRNA libraries are enriched with piRNAs. Therefore, to compare piRNA abundances across different oxidized libraries, we calibrated oxidized to unoxidized libraries. Because paired oxidized and unoxidized sRNA libraries were created from the same source, the subset of piRNA species should remain constant between the two libraries. First, unoxidized libraries were normalized to sequencing depth (ppm; parts per million). Next, we identified all the uniquely mapping piRNA species (piRNAs that shared the exact nucleotide sequences) that were shared between at least two of the three replicates of paired oxidized and unoxidized libraries. Finally, the calibration factor

was computed using the ratio between the sums of the normalized abundance in the unoxidized libraries and the abundances in the oxidized libraries:

$$calibration\_factor = \frac{\sum_c ppm_{c,unox}}{\sum_c counts_{c,ox}}$$

Where *c* is the number of common piRNA species between oxidized and unoxidized libraries. To avoid potential bias caused by outliers, we did not include the piRNAs in the top 10th percentile in the calculation. The number of copies for each piRNA in the oxidized library was finally calculated by multiplying by the calibration factor.

RNA-seq Library Preparation and Analysis

RNA-seq libraries were constructed as described in (Zhang et al., 2012) with a few modifications (Wu, Fu, and Zamore, manuscript in preparation). For ribosomal RNA depletion, RNA was hybridized in 10 µl with a pool of 186 rRNA antisense oligos (0.05µM/each; Morlan et al., 2012; Adiconis et al., 2013) with 10 mM Tris-HCl [pH 7.4] and 20 mM NaCl and heated to 95°C, then cooled at -0.1°C/sec to 22°C, and finally incubated at 22°C for 5 min. Ten units of RNase H (Lucigen, #H39500) were added and incubated at 45°C for 30 min in 20 µl containing 50 mM Tris-HCl [pH 7.4], 100 mM NaCl, and 20 mM $MgCl_2$. RNA was then treated with 4 units DNase (Thermo Fisher, #AM2238) in 50 µl at 37°C for 20 min. After DNase treatment, RNA was purified using RNA Clean & Concentrator-5 (Zymo Research, #R1016). RNA-seq libraries were sequenced using a NextSeq500 (Illumina) to obtain 75 + 75 nt, paired-end reads.

Barcodes were sorted allowing one mismatch, and the 3′ adaptors, including the three random nucleotides, were identified and removed using the first ten nucleotides, allowing one mismatch. RNA-seq analysis was performed with piPipes (v1.4; Han et al., 2014). Briefly, RNAs were first aligned to rRNA sequences using Bowtie2 (v2.2.0; Langmead and Salzberg, 2012). Unaligned reads were then mapped using STAR to the fly genome (v2.3.1; Dobin et al., 2013). Counts were produced using the “strict” option on HTseq (v0.6.1; Anders et al., 2015).

ChIP-seq Library Preparation

ChIP-seq libraries were constructed as described in (Zhang et al., 2014) with a few modifications. Briefly, ~100 µl ovaries per library were first crosslinked with 2% formaldehyde for 10 min rotating at 25°C in Robb's medium (100 mM HEPES [pH 7.4], 55 mM sodium acetate, 40 mM potassium acetate, 100 mM sucrose, 10 mM glucose, 1.2 mM $MgCl_2$, 1 mM $CaCl_2$, 1mM DTT, 1 mM AEBSF, 0.3 µM Aprotinin, 20 µM Bestatin, 10 µM E-64, and 10 µM Leupeptin). Crosslinking was quenched by adding Glycine to a final concentration of 120 mM and for 5 min rotating at 25°C. The ovaries were then washed twice with TBS (50 mM Tris-HCl [pH 7.5], 150 mM NaCl), and twice with ChIP lysis buffer (50 mM Hepes/KOH [pH 7.5], 140 mM NaCl, 1% [v/v] Triton X-100, 0.1% [w/v] Na-Deoxycholate, 0.1% [w/v] SDS).

Ovaries were then sonicated in sonication buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl [pH 8.0], 1mM DTT, 1 mM AEBSF, 0.3 µM Aprotinin, 20 µM Bestatin, 10 µM E-64, and 10 µM Leupeptin) using an E220 Evolution Focused-ultrasonicator (Covaris). The following sonication parameters were used:

Duty cycle: 5%

Intensity: 140 watts

Cycles per burst: 200

Temperature: <10°C

Time: 20 min

The sonicated lysate was centrifuged at 13,000 × *g* for 15 min at 4°C. Supernatant was diluted 7-fold with dilution buffer (20 mM Tris-HCl [pH 7.5], 167 mM NaCl, 1.2 mM EDTA, 0.01% [w/v] SDS, 1% [v/v] Triton X-100, 1mM DTT, 1 mM AEBSF, 0.3 µM Aprotinin, 20 µM Bestatin, 10 µM E-64, and 10 µM Leupeptin) and incubated overnight rotating at 4°C with antibody (anti-Rhi or Pre-Immune Serum, gift from William Theurkauf, 20 µl; anti-HP1a, DSHB, #C1A9, 5 µg; normal mouse IgG, Abcam,

#ab188776, 5 µg; anti-H3K9me3, Abcam, #ab8898, 10.5 µg; anti-H3K4me3, Abcam, #ab8580, 10.5 µg; anti-Histone H3, Abcam #ab18521, 10.55 µg) conjugated to 100 µl of Dynabeads Protein A/G (Life Technologies, #10002D/#10004D).

The beads were then washed 2×5 min each with 500 µl of the following buffers: Wash buffer A (20 mM Tris-HCl [pH 8.0], 2 mM EDTA, 0.1% [w/v] SDS, 1%[v/v] Triton X-100, 150 mM NaCl), Wash buffer B (20 mM Tris-HCl [pH 8.0], 2 mM EDTA, 0.1% [w/v] SDS, 1%[v/v] Triton X-100, 500 mM NaCl), Wash buffer C (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1% [v/v] NP-40, 1% [w/v] Na-deoxycolate, 0.25 M LiCl) and Wash buffer D (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). All wash buffers also contained 1mM DTT, 1 mM AEBSF, 0.3 µM Aprotinin, 20 µM Bestatin, 10 µM E-64, and 10 µM Leupeptin. Beads were then treated with 20 µg/ml RNase A (Fisher Scientific, #FEREN0531) To reverse crosslink and remove protein, beads were incubated overnight at 65°C with 200 µg/ml Proteinase K (Life Technologies, #25530015) in 2×Proteinase K Buffer (200 mM Tris-HCl [pH 7.5], 2 mM EDTA [pH 8.0], and 1% SDS (w/v). Finally, DNA was purified using phenol:chloroform [pH 8.0] and the library was prepared by sequentially performing end-repair, A-tailing, Y-shaped adaptor ligation, and PCR amplification as described in in (Zhang et al., 2012).

Analysis of H3K9me3, HP1a, and Rhi ChIP-seq Libraries

Barcodes were sorted allowing one mismatch, and the 3′ adaptors, including the three random nucleotides, were identified and removed using the first ten nucleotides, allowing one mismatch. Reads were mapped on the genome using Bowtie2 (v2.2.0; Langmead and Salzberg, 2012). Unmapped reads were removed using SAMtools (v0.1.19; Li et al., 2009b; Li, 2011) and a mapping *q*-score of 10 was used to identify uniquely mapping reads. We used a 1 kbp sliding window with a 500 bp step over the genome to compute a signal for each chromosome. For each chromosome, the counts

for each bin were normalized using the total number of reads mapping to the chromosome.

Unique reads were also mapped to genomic features (genes, transposons, and piRNA producing loci) using STAR (v2.3.1; Dobin et al., 2013) and counts were produced using the "strict" option on HTseq (v0.6.1; Anders et al., 2015). Reads were normalized to sequencing depth.

Analysis of H3K4me3 ChIP-seq Libraries

psychENCODE pipeline was used to analyze H3K4me3 ChIP-seq libraries. Briefly, reads were mapped to the genome using BWA (Li and Durbin, 2009). SAMtools (v0.1.19; Li et al., 2009b; Li, 2011) and Picard Toolkit (http://broadinstitute.github.io/picard/Ioannidis, 2005) were used to remove improperly paired reads and PCR duplicates. MACS2 (Zhang et al., 2008b) was used to called peaks. We used BEDtools (v2.26.0; Quinlan and Hall, 2010) to merge peaks from all replicates for each genotype to create a consensus set of peaks. The number of reads overlapping each peak was computed using BEDtools and reads were normalized to sequencing depth.

GRO-seq Library Preparation and Analysis

GRO-seq was performed as described in (Wang et al., 2015) with a few modifications. Briefly, 0–2-day-old female flies were given yeast for 2 days before their ovaries were dissected. One hundred pairs of ovaries were homogenized in 350 µl HB35 buffer (15 mM HEPES KOH [pH 7.5], 10 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 0.05% [v/v] NP 40, 0.35 M sucrose, 1 mM DTT, 1 mM AEBSF, 0.3 µM Aprotinin, 20 µM Bestatin, 10 µM E-64, and 10 µM Leupeptin) with a Dounce homogenizer using pestle B (Sigma Aldrich, #D8938). Nuclei were purified by passing twice through sucrose cushions that contain 800 µL HB80 buffer (15 mM HEPES KOH [pH 7.5], 10 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 0.05% [v/v] NP 40, 0.80 M sucrose, 1

mM DTT, 1 mM AEBSF, 0.3 μM Aprotinin, 20 μM Bestatin, 10 μM E-64, and 10 μM Leupeptin) on the bottom phase and 350 μl HB35 buffer on the top. Nuclei were washed once with 500 μl freezing buffer (50 mM Tris-HCl, pH 8.0, 40% [v/v] glycerol, 5 mM MgCl2, 0.1 mM EDTA, 1 mM dithiothreitol, 1 mM AEBSF, 0.3 μM Aprotinin, 20 μM Bestatin, 10 μM E-64, and 10 μM Leupeptin) and frozen in liquid nitrogen with 100 μL freezing buffer. To carry out nuclear run on assay, 100 μL freshly prepared reaction buffer (10 mM Tris-HCl, pH 8.0, 5 mM MgCl2, 300 mM KCl, 1% [w/v] sarkosyl, 500 μM ATP, 500 μM GTP, 500 μM Br UTP, 2.3 μM CTP, 1 mM dithiothreitol, 20 U RNasin Plus RNase Inhibitor (Promega, #N2615) was added to nuclei and incubated at 30°C for 5 min. RNA was extracted using Trizol (Invitrogen, #15596). Nascent RNAs with Br UTP incorporated were enriched by immunoprecipitation using anti 5 bromo 2′ deoxyuridine antibody (Fisher Scientific, #50175223) as described (Shpiz and Kalmykova, 2014), followed by rRNA depletion using RNase H, fragmentation, and library construction as in RNA-seq library preparation (Zhang et al., 2012). Analysis was carried out using the RNA seq pipeline in piPipes.

Ping-pong Analysis

Ping-pong analysis was conducted as in (Zhang et al., 2011). In summary, for two piRNAs that were sufficiently complementary to each other at a particular 5′-to-5′ distance, a *Z* score was defined as the product of their abundances. The Ping-Pong Z_{10} score was then the difference of the score at the 5′-to-5′ distance of 10 nt and the mean scores of background distances, divided by the standard deviation of the scores of background distances, defined as distances of 0–9 and 11–20 nt. Two piRNAs were sufficiently complementary to each other when the nucleotides 2–10 of the first piRNA were perfectly paired with the second piRNA and there was at most one mismatch among positions 1 and 11–16 of the first piRNA. Genomic sequence adjacent to the second piRNA was used to determine complementarity when the 5′-to-5′ distance was

less than 16 nt. For analyses including multi-mappers, reads were apportioned by the number of times they aligned to the genome.

Phasing Analysis

Phasing analysis was done as in (Han et al., 2015). Briefly, sRNA reads were mapped to genome and rRNAs, tRNAs, and snoRNAs were removed. The Z_x score for a distance x between the 3′ end of one piRNA to the 5′ end of a downstream piRNA on the same genomic strand was calculated by the difference of the score at the distance x and the mean scores of background distances, divided by the standard deviation of the scores of background distances. When x = 1, the 5′ end is immediately downstream of the 3′ end (phasing). For analyses including multi-mappers, reads were apportioned by the number of times they aligned to the genome. To calculate Z_1, overlapping nts at position 2-20 were used as background to calculate Z scores.

Immunohistochemistry and Microscopy

Immunohistochemistry and microscopy was performed as described in (Wang et al., 2015). Anti-Rhi (1:100 dilution) was a gift from William Theurkauf (Zhang et al., 2014). Images presented in the same figure were acquired at the same settings.

Female Fertility Assay

Female fertility was tested as described in (Li et al., 2009a) with a few changes: Eight female virgins were mated to four Oregon R virgin males in a small cage with a 60 mm diameter grape juice agar plate dabbed with yeast paste at 25°C. At collection, all flies were collected <1-day post-eclosion. After two days, the first plate was discarded and replaced with a fresh plate. Plates were then changed and scored every subsequent day. The number of total eggs, eggs per female per day, and the dorsal appendage phenotype of embryos were scored every 24 h and the number of eggs that hatched

were scored 48 h after the plate was changed. Fertility was recorded for 12 days and at least two independent biological replicates were conducted for each genotype.

Eye Pigment Assay

Ethanol-based pigment extraction and quantification was performed as described in (Sun et al., 2004) with a few changes. Briefly, 10 females, 3–5 days post-eclosion, were collected and their heads photographed and/or dissected. Heads were homogenized with a plastic pestle (Fisher Scientific, #12141364) in 0.5 ml of 0.01 M HCl in ethanol. The homogenate was incubated at 4°C rotating overnight, warmed to 50°C for 5 min, and centrifuged at 13,000 × *g* for 10 min at 25°C. The supernatant was collected and the optical density at 480 nm (OD480) was measured. To accurately measure only the eye pigment, the OD480 for 10 female w^{1118} heads was subtracted from each measurement. Three independent biological replicates were collected for each genotype. Mean values and standard deviation were compared between matching $mael^{M391/r20}$ and control flies. *p*-values were measured using an unpaired, two-tailed t-test.

ACKNOWLEDGEMENTS

We thank William Theurkauf for the Ago3 antibody; Julius Brennecke for the Mael antibody; Toshie Kai for the FLAG-Mael rescue fly; Cindy Tipping and Alicia Boucher for fly husbandry, and members of the Weng and Zamore laboratories for help, advice, discussions, and comments on the manuscript. This work was supported in part by National Institutes of Health grants P01HD078253 to Z.W. and P.D.Z. and GM65236 to P.D.Z.

CHAPTER III: Discussion

Mael Keeps Rhi in Check

Heterochromatin has long been considered repressive and inert. Therefore, it is not a surprise that *D. melanogaster* piRNA clusters—which are composed of transposons—are heterochromatic: transposon sequences are dangerous and need to be silenced. H3K9me3 and HP1a alone, however, are not sufficient to repress all transposons, as some transposons are active in the absence of Mael or Piwi despite little change in heterochromatin (Sienski et al., 2012; Klenov et al., 2014; Figures 2.5 and 2.7B). Furthermore, recent studies revealed that in the fly germline, dual-strand piRNA clusters are heterochromatic not just to repress coherent, promoter-initiated transcription, but to allow incoherent transcription mediated by Rhi (Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014; Andersen et al., 2017). Like other HP1 proteins, Rhi acts like a scaffold protein and mediates the interaction of diverse proteins to couple non-canonical transcription of dual-strand piRNA clusters and the export of piRNA precursor transcripts to the cytoplasm to be processed into mature piRNAs (Zhang et al., 2012; Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014; Chen et al., 2016; Hur et al., 2016; Andersen et al., 2017). Thus, Rhi allows flies to bypass the paradoxical problem that transposon sequences must be transcribed to create piRNAs that are then used to silence the same sequences from which they were created.

If Rhi is required for sequence-independent transcription initiation, then how does it also promote canonical transcription in dual-strand clusters? One possibility is that transcription, whether canonical or non-canonical, changes chromatin organization and allows promoters usually hidden by heterochromatin to be accessible to transcription factors (Mavrich et al., 2008; Parnell et al., 2008; Schones et al., 2008; Vermaak and Malik, 2009). Thus, Mael is required to prevent coherent, promoter-driven, transcription in dual-strand clusters that is enhanced by Rhi. Indeed, more H3K4me3 peaks were detected throughout piRNA clusters in *mael* mutant ovaries. While H3K4me3 is not an

activating signal per se, it is highly correlated with transcription start sites of canonically transcribed genes (Howe et al., 2017). Furthermore, the loss of Rhi suppressed canonical transcription improperly activated in dual-strand clusters of *mael* mutant ovaries (Figure 2.18). Therefore, our data are consistent with a model that Rhi is required to maintain a permissive transcriptional environment in dual-strand piRNA clusters that is amenable to both canonical and non-canonical transcription. Previously, the Brennecke lab proposed that Mael functions downstream of Piwi and H3K9me3 to transcriptionally repress transposons (Sienski et al., 2012). We would like to update this model and propose that Mael, guided by Piwi, is required to repress coherent, promoter-dependent transcription that is incidentally activated by Rhi (Figure 3.1).

Figure 3.1

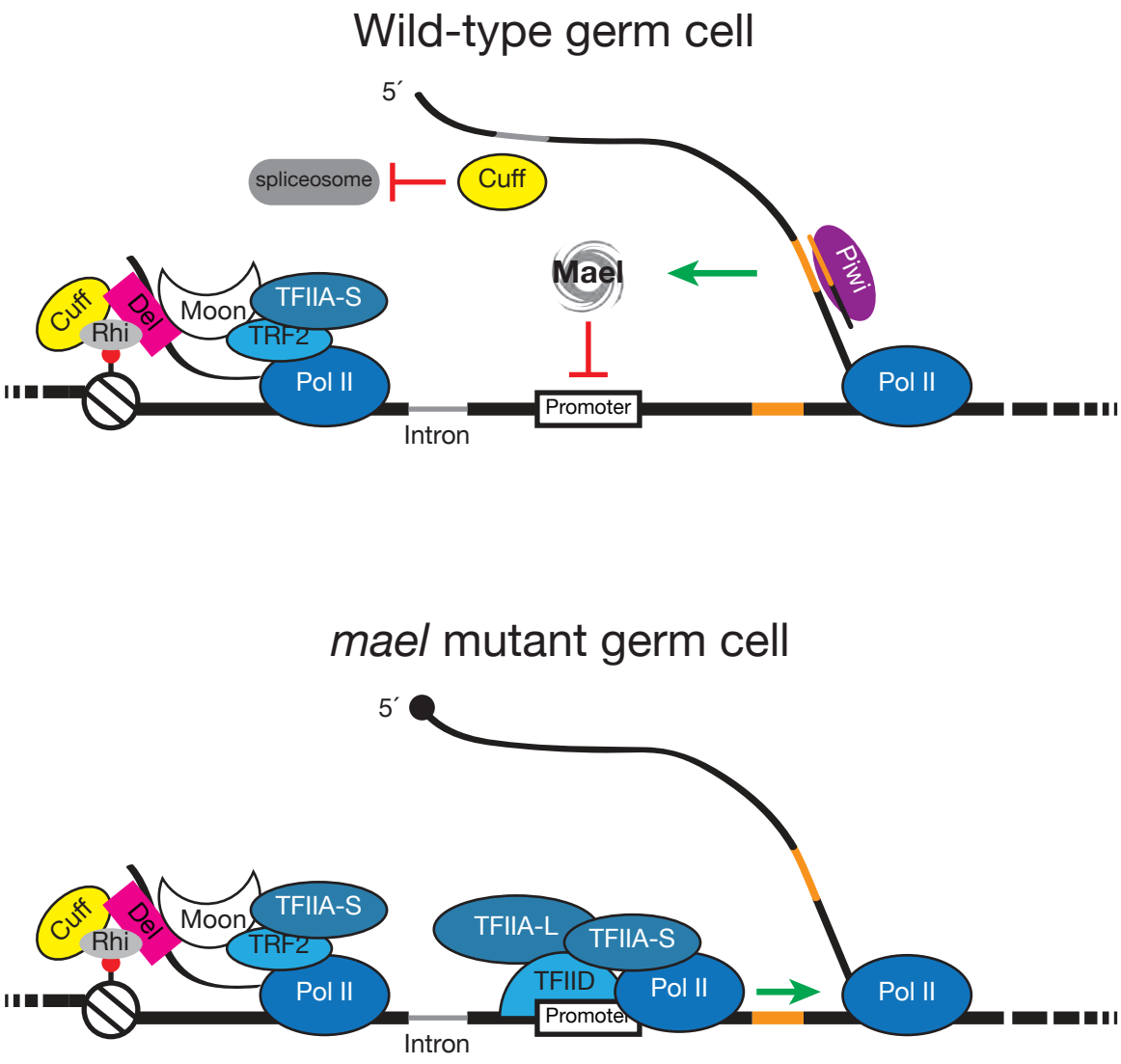

Wild-type germ cell
5´
spliceosome
Cuff
Mael
Piwi
Cuff
Del
Rhi
Moon
TFIIA-S
TRF2
Pol II
Promoter
Pol II
Intron
mael mutant germ cell
5´
Cuff
Del
Rhi
Moon
TFIIA-S
TRF2
Pol II
TFIIA-L
TFIIA-S
TFIID
Pol II
Promoter
Pol II
Intron

Figure 3.1: A model for Mael-dependent repression of canonical transcription in dual-strand piRNA clusters

In *D. melanogaster* germ cells, dual-strand piRNA cluster transcription is dependent on Rhi and associated proteins, which allow Pol II to initiate independent of promoters. One Rhi-associated protein, Cuff, stabilizes uncapped transcripts and suppresses splicing. While “incoherent,” Rhi-mediated transcription may expose promoters that would otherwise be hidden by heterochromatin. Piwi, guided by piRNAs, binds to nascent transcripts and targets different proteins to genomic loci to repress transcription. One of these proteins is Mael, which blocks promoter-driven transcription. Without Mael, canonical transcripts are produced from dual-strand clusters which can be translated into protein.

Mael-mediated transcription repression appears to only affect canonical transcription. Considering that wild-type ovaries do not show consistent Pol II ChIP-seq signal from dual-strand clusters (Mohn et al., 2014), and have few RNA- or GRO-seq reads from *P{GSV6}42A18* with or without GAL4-VP16 (Figure 2.4), it seems likely that Mael represses transcription at an early step, such as formation of the pre-initiation complex (PIC). Because Moon forms an alternative TFIIA-TBP complex in ovaries consisting of TFIIA-S and TRF2 (Andersen et al., 2017), Mael might prevent formation of the PIC by inhibiting the binding of transcription factors and/or general transcription factors, such as TFIID/TBP, from binding to promoters. Our attempts to test the access of transcription factors to the *P{GSV6}42A18* promoter by immunoprecipitating GAL4-VP16 have been unsuccessful. Looking into the interaction of general transcription factors and dual-strand clusters could reveal how non-canonical transcription is permitted yet canonical transcription is inhibited.

According to our model, a potential paradox exists when we compare $rhi^{2/KG}$ single mutants to $rhi^{2/KG}; mael^{M391/r20}$ double mutant ovaries (Figures 2.17 and 2.18): *if Mael is guided to its targets by Piwi-loaded piRNAs, then how does Mael repress canonical transcription in dual-strand piRNA clusters in* $rhi^{2/KG}$ *ovaries?* Because embryonic Piwi is required to license dual-strand clusters (Akkouche et al., 2017), it is possible that maternally deposited Piwi also guides Mael to its targets early in development. Consistent with this hypothesis, germline knockdown of Mael, which depletes Mael in adult ovaries but does not affect maternally deposited Mael, did not cause the derepression of *gfp* (Figure 2.4A). These results suggest that Mael is only required in the early embryo to repress canonical transcription and therefore must leave a persistent signal at dual-strand clusters. The simplest explanation would be that Mael helps modify chromatin that is maintained throughout development. While Mael does not influence H3K9me3, HP1a, or Rhi, there are other possible chromatin modifications that could be altered in *mael* mutants. One caveat with knockdown experiments is that

low levels of Mael may be sufficient to repress canonical transcription. Furthermore, Mael germline depletion still led to transposon derepression and sterility (Czech et al., 2013). Therefore, Mael may have multiple roles at different developmental stages. We will explore this possibility in greater depth in the next section.

In summary, it appears that while Rhi allows flies to transcribe heterochromatic transposon sequences in dual-strand clusters, it is not without its hazards. Therefore, Mael is required to prevent incidental canonical, promoter-dependent transcription made possible by Rhi.

Mael is Required for Dual-Strand piRNAs

Although Mael is downstream of Piwi, *mael*$^{M391/r20}$ ovaries have fewer piRNAs than control (Figure 2.10). In particular, piRNAs mapping to dual-strand clusters, including *P{GSV6}42A18*, were most affected. While the Brennecke lab also noted the loss of *42AB* piRNAs in *mael*$^{M391/Df1}$ mutant ovaries the loss of piRNAs was less severe than what we detected in *mael*$^{M391/r20}$ mutant ovaries (Sienski et al., 2012). One possible explanation for this discrepancy could be differences in the fly strains used. While both studies used the *mael*^{M391} allele, the previous study used *mael*^{Df1}, a deficiency which removes ~400 kbp from chromosome 3L. Other than *mael*, no known piRNA pathway genes are found in this region, although there are a few miRNAs, unannotated genes, and lncRNAs that are lost in *mael*^{Df1}. Furthermore, both *mael*^{M391} and *mael*^{r20} were outcrossed for five generations for this study and therefore *mael*$^{M391/r20}$ and *mael*$^{M391/Df1}$ mutant flies may have different background mutations and transposon copy numbers. Finally, we made paired oxidized and unoxidized sRNA libraries from six independent *mael*$^{M391/r20}$ mutant replicates with different transgene insertions. While there are some differences across replicates, overall, *mael*$^{M391/r20}$ ovaries have fewer piRNAs than control.

At first glance, the cause of the loss of piRNAs in *mael* mutants was not obvious: both ping-pong and phasing appeared functional yet both processes occurred less in *mael*$^{M391/r20}$ ovaries (Figures 2.11 and 2.13). The decrease in ping-pong pairs was likely partially caused by the loss of *ago3* transcripts in *mael*$^{M391/r20}$ mutant ovaries (Figure 2.12). The loss of Ago3 is unlikely to be the sole cause of the decrease in piRNAs, however, because *piwi*$^{2/Nt}$ ovaries express similar levels of *ago3* mRNA yet have normal levels of cluster piRNAs (Table 2.1 and Figure 2.14B).

Another possibility for the loss of piRNAs in *mael*$^{M391/r20}$ mutants is that Rhi-mediated transcription initiation and export of cluster transcripts become decoupled by canonical transcription. Although canonical and non-canonical transcription are mutually exclusive at a specific locus, both are likely occurring at separate loci across dual-strand clusters in *mael*$^{M391/r20}$ mutant ovaries. This is supported by our findings that while there were fewer piRNAs and ping-pong is rarer, both were not completely lost without Mael. Furthermore, because canonical transcripts are capped, spliced, and polyadenylated, Cuff, which is thought to recognize the 5′ ends of piRNA precursors (Mohn et al., 2014; Zhang et al., 2014; Chen et al., 2016), may not be able to interact with canonical, coherent transcripts produced in *mael*$^{M391/r20}$ ovaries. Without Cuff, two proteins required for export of piRNA precursors to the nuage, UAP56 and Thoc5, are not loaded and are no longer localized to the nucleus (Zhang et al., 2012; Hur et al., 2016; Zhang et al., 2014). Thus, rather than getting processed into piRNAs, coherent transcripts are translated into protein (Figure 2.4).

Finally, Mael may help transport dual-strand cluster transcripts to the nuage to be processed into mature piRNAs. According to this model, fewer piRNA precursors are available in the nuage to be processed into piRNAs in *mael*$^{M391/r20}$ ovaries; explaining why we detected less ping-pong, phasing, and piRNAs. While *mael*$^{M391/r20}$ mutants had more potential piRNA precursors (Figures 2.4, 2.5, and 2.6), to test this hypothesis, more experiments are needed to determine whether piRNA precursors are trapped in

the nucleus of *mael*$^{M391/r20}$ germ cells. Furthermore, consistent with this model, Mael was previously shown to be able to shuttle between the nucleus and the cytoplasm (Findley, 2003). In addition, the MAEL domain is predicted to be an RNA-binding domain (Zhang et al., 2008a) and interacted with ssRNA in vitro (Chen et al., 2015; Matsumoto et al., 2015). Interestingly, although Mael lacks classical nuclear localization or export signals, point mutations to the MAEL domain caused Mael to lose both nuclear and nuage localization and were unable to rescue transposon repression or fertility (Sienski et al., 2012). Finally, mouse MAEL immunoprecipitated with piRNA precursors and was also required for normal piRNA production (Castaneda et al., 2014). While our attempts at immunoprecipitating Mael and sequencing RNAs associated with it (RNA-IP) were unsuccessful, we cannot rule out the possibility that Mael participates in RNA transport or early in piRNA processing.

Although our data support Mael repressing transcription downstream of Piwi, due to the cyclical nature of the piRNA pathway, it can be difficult to discern the exact role and hierarchy for different factors through genetics. Therefore, it is possible that because *mael*$^{M391/r20}$ mutant ovaries have fewer piRNAs, Piwi cannot effectively repress transposons in the nuclei.

Mael Also Represses Transcription in the Soma

Our data are consistent with previous reports that Mael functions downstream of H3K9me3 to repress transposons in OSCs (Sienski et al., 2012). Similar to germ cells, in somatic follicle cells, Mael repressed canonical transcription in *42AB* (Figure 2.20B). Furthermore, we found that Mael was strong suppressor of PEV (Figure 2.21 and Table 2.3). While Mael or Piwi is not readily detectable outside the ovaries, depleting Piwi in early embryos suppressed the variegation of a somatic reporter in adult flies (Gu and Elgin, 2013). Therefore, in early embryos, Piwi was required for establishing dual-strand clusters in the germline and heterochromatin in the soma (Gu and Elgin, 2013;

Akkouche et al., 2017). Similar to the germline, Piwi may also guide Mael to somatic targets early in development. Moreover, because Armi, Piwi, and Mael are all expressed in somatic follicle cells, Piwi may actively guide Mael to repress transcription of heterochromatic sequences in these cells.

In w^{m4} flies, the breakpoints that caused the chromosomal inversion of *white* are transposons. Furthermore, in w^{m4} revertant lines, in which *white* and nearby transposon sequences were reinverted back into euchromatin, the transposon sequences were still heterochromatic (Tartof et al., 1984). This suggests that heterochromatin was actively deposited at these transposon sequences and did not just propagate from the centromere. Therefore, in w^{m4}; $mael^{M391/r20}$ mutant eyes, it is possible that the derepression of transposons near *white* also caused an increase in *white* expression, thereby suppressing PEV. Historically, suppressors of PEV were often involved with the silencing properties of heterochromatin. Our data suggest that in flies, Mael represses transcription of heterochromatic sequences inside and outside the gonads.

A Conserved Role for Mael

A role for Mael outside the adult gonads that may or may not be piRNA-dependent is not surprising because the MAEL domain is conserved in protists, which have neither gonads nor piRNAs (Zhang et al., 2008a). In protists, the MAEL domain is an active nuclease and it is interesting to speculate that Mael might have a conserved role repressing unwanted transcripts (Chen et al., 2015). In mice, while the MAEL domain has lost its catalytic residues, it associates with modified P-bodies, distinct cytoplasmic loci consisting of many enzymes involved in mRNA turnover, and thus may also be involved with post-transcriptional regulation (Zhang et al., 2008a; Aravin et al., 2009; Kulkarni et al., 2010).

Although *D. melanogaster* does not have somatic piRNAs, most arthropods spanning Chelicerata (*Parasteatoda tepidariorum*; common house spider), Myriapoda

(*Strigamia maritima*; coastal European centipede), and Insecta (including *Drosophila virillis*, fruit fly; *Aedes aegypti*, yellow fever mosquito; and *Trichoplusia ni* [*T. ni*], cabbage looper) have both somatic piRNAs and Mael homologs (Zhang et al., 2008a; Fu et al., 2018b; Lewis et al., 2018). If these arthropods also express somatic Mael, PIWI proteins may actively guide Mael to repress transposons outside the gonads.

Although *T. ni* express somatic piRNAs, it is an especially attractive model organism to study germline piRNAs because the Hi5 cell line, derived from *T. ni* ovarian germ cells, expresses two PIWI proteins and produces piRNAs that show significant ping-pong signatures (Granados et al., 1986; Fu et al., 2018b). Furthermore, because cabbage loopers do not have Rhi, Del, or Cuff orthologs, all of which are poorly conserved, *T. ni* may provide a more universal insect model for piRNA cluster transcription (Fu et al., 2018b). Finally, the dipteran, *D. melanogaster*, and the lepidopteran, *T. ni*, diverged over 300 million years ago (Reuter et al., 2009). Therefore, Hi5 cells would make an ideal system to test whether transcription repression by Mael is conserved in other insects.

In addition to the loss of piRNAs, *mael* mutant mice, like flies, are also sterile and cannot repress germline transposons (Costa et al., 2006; Soper et al., 2008; Aravin et al., 2009; Castaneda et al., 2014). Furthermore, in both flies and mice, Mael does not play a major role in heterochromatin formation, as the loss of mouse MAEL did not affect DNA methylation of L1 elements (Aravin et al., 2009). Despite these similarities, there are some differences as well. While fly Mael does not appear to interact with other piRNA pathway factors (Sato et al., 2011), mouse MAEL colocalizes and associates with several RNA metabolism factors and piRNA pathway components in the cytoplasm including piRNA cluster transcripts and transposons (Aravin et al., 2009; Castaneda et al., 2014). Furthermore, mouse MAEL has yet to be linked to transcriptional repression and current evidence from mouse suggests that MAEL is involved with the processing of piRNA precursors (Soper et al., 2008; Aravin et al., 2009; Castaneda et al., 2014).

Although a cytoplasmic role for Mael in the piRNA pathway has yet to be defined in flies, in both organisms, Mael prevents transposon sequences normally repressed by heterochromatin from becoming expressed.

Conclusions: "There is Nothing New Under the Sun"

Life on Earth is astonishingly diverse, yet all organisms still face a common set of problems. While evolution may take different paths to reach a solution, we see similar motifs used over and over. One such theme is how species maintain their genomes and protect themselves from dangerous and mutagenic elements. In this thesis, I hoped to have expanded the knowledge of how *D. melanogaster* represses transposons using a variation of the common trope: transcriptional repression through sRNAs. While sRNAs are extremely flexible, they must be made from sequences that are similar to those that must be silenced. Thus, another common problem many organisms face is how to silence a sequence when the transcription of the sequence is required for silencing?

Like *A. thaliana* and *S. pombe*, *D. melanogaster* uses heterochromatin to transcriptionally repress potentially harmful DNA sequences. To bypass heterochromatic silencing, *A. thaliana* uses special polymerases, *S. pombe* transcribes these sequences at specific times in the cell cycle, and *D. melanogaster* can initiate transcription independent of promoters. Because transcripts from these sequences are potentially dangerous, they are turned into dsRNA in *A. thaliana* and *S. pombe* and are incoherent in *D. melanogaster*. As an additional safeguard, all three organisms couple transcription of these sequences to sRNA processing. Finally, a common theme in sRNA-mediated transcriptional repression is that it is a self-reinforcing loop with the ability to spread to neighboring DNA. After sRNAs find their targets, they can initiate heterochromatin formation which then brings additional components to produce more sRNAs. Thus, despite the differences between these three organisms spanning across three kingdoms of life, they have all evolved parallel methods to repress transcription.

While the end results may be similar, “the devil is in the detail,” and evolution has taken fascinating paths in each of these individual mechanisms. In *D. melanogaster*, one such “detail” is the enigmatic protein, Maelstrom. In this dissertation, we showed that Mael is required to repress canonical transcription in dual-strand piRNA clusters. Without Mael, promoters inside dual-strand clusters become activated and potentially dangerous transposons are transcribed. We also expand upon the finding that heterochromatin is not always silent and provide additional details into how transcription in heterochromatin is regulated.

Furthermore, we found that Mael is required to produce dual-strand piRNAs. Despite the increase in cluster transcription, without Mael, canonical transcripts are not processed into piRNAs, hinting at a possible selective mechanism by which germ cells differentiate piRNA precursors from endogenous mRNAs.

Finally, using fly strains from Muller’s original PEV lines and newer tools to manipulate the genome, we showed that Mael also represses canonical transcription outside the germline and provided new insight into how heterochromatic domains are silenced in somatic tissue.

While the destructive potential of transposons must be contained, transposons also play a uniquely generative role in genome evolution as well. Perhaps epigenetic mechanisms, such as the piRNA pathway, may not have evolved to repress transposons, but rather to allow the accumulation of transposons (Fedoroff, 2012). The fingerprints of transposons can be seen throughout the eukaryotic genome, ranging from co-opted genes to the expansion of the eukaryotic genome. Thus, Maelstrom may actually be a guiding force in the evolution of our genome.

CHAPTER IV: BIBLIOGRAPHY

Aasland, R., and Stewart, A. F. (1995). The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. Nucleic Acids Res *23*, 3168-3173.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K.,

Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., WoodageT, Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., and Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., Sivachenko, A., Thompson, D. A., Wysoker, A., Fennell, T., Gnirke, A., Pochet, N., Regev, A., and Levin, J. Z. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods *10*, 623-629.

Akkouche, A., Mugat, B., Barckmann, B., Varela-Chavez, C., Li, B., Raffel, R., Pélisson, A., and Chambeyron, S. (2017). Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. Mol Cell

Allshire, R. C., Javerzat, J. P., Redhead, N. J., and Cranston, G. (1994). Position effect variegation at fission yeast centromeres. Cell *76*, 157-169.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andersen, P. R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature

Arabidopsis, G. I. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., and Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A. A., Klenov, M. S., Vagin, V. V., Bantignies, F., Cavalli, G., and Gvozdev, V. A. (2004). Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. Mol Cell Biol *24*, 6742-6750.

Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., and Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol *11*, 1017-1027.

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., and Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Aravin, A. A., van der Heijden, G. W., Castaneda, J., Vagin, V. V., Hannon, G. J., and Bortvin, A. (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. PLoS Genet *5*, e1000764.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Baulcombe, D. C. (1999). RNA makes RNA makes no protein. Curr Biol *9*, R599-R601.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol *14*, 336-342.

Bentley, D. L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol *17*, 251-256.

Bergman, C. M., Quesneville, H., Anxolabéhère, D., and Ashburner, M. (2006). Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol *7*, R112.

Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A *99*, 8695-8700.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bianchi, M. E., and Agresti, A. (2005). HMG proteins: dynamic players in gene regulation and differentiation. Curr Opin Genet Dev *15*, 496-506.

Bingham, P. M., Kidwell, M. G., and Rubin, G. M. (1982). The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell *29*, 995-1004.

Blair, R. H., Goodrich, J. A., and Kugel, J. F. (2012). Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics. Biochemistry *51*, 7444-7455.

Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., Tang, H., and Pikaard, C. S. (2015). Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. Elife *4*, e09591.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., and Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387-1392.

Bühler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873-886.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.

Cam, H. P., Sugiyama, T., Chen, E. S., Chen, X., FitzGerald, P. C., and Grewal, S. I. (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat Genet *37*, 809-819.

Carmell, M. A., Girard, A., van de Kant, H. J., Bourc'his, D., Bestor, T. H., de Rooij, D. G., and Hannon, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Castaneda, J., Genzor, P., van der Heijden, G. W., Sarkeshik, A., Yates, J. R., Ingolia, N. T., and Bortvin, A. (2014). Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. EMBO J

Cenik, E. S., and Zamore, P. D. (2011). Argonaute proteins. Curr Biol *21*, R446-9.

Cerutti, L., Mian, N., and Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends Biochem Sci *25*, 481-482.

Chalvet, F., Teysset, L., Terzian, C., Prud'homme, N., Santamaria, P., Bucheton, A., and Pélisson, A. (1999). Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. EMBO J *18*, 2659-2669.

Chang, J. H., Jiao, X., Chiba, K., Oh, C., Martin, C. E., Kiledjian, M., and Tong, L. (2012). Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity. Nat Struct Mol Biol *19*, 1011-1017.

Chen, E. S., Zhang, K., Nicolas, E., Cam, H. P., Zofall, M., and Grewal, S. I. (2008). Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature *451,* 734-737.

Chen, K. M., Campbell, E., Pandey, R. R., Yang, Z., McCarthy, A. A., and Pillai, R. S. (2015). Metazoan Maelstrom is an RNA-binding protein that has evolved from an ancient nuclease active in protists. RNA

Chen, Y., Pane, A., and Schupbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. Curr Biol *17,* 637-642.

Chen, Y. A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaya, E., Li, S., Vempati, S., Laver, J. D., Patel, D. J., Smibert, C. A., Lipshitz, H. D., Fejes Toth, K., and Aravin, A. A. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. Mol Cell

Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436,* 740-744.

Cheutin, T., McNairn, A. J., Jenuwein, T., Gilbert, D. M., Singh, P. B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science *299,* 721-725.

Chung, W. J., Okamura, K., Martin, R., and Lai, E. C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr Biol *18,* 795-802.

Clarke, L., and Baum, M. P. (1990). Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol Cell Biol *10,* 1863-1872.

Clegg, N. J., Findley, S. D., Mahowald, A. P., and Ruohola-Baker, H. (2001). Maelstrom is required to position the MTOC in stage 2-6 Drosophila oocytes. Dev Genes Evol *211,* 44-48.

Clegg, N. J., Frost, D. M., Larkin, M. K., Subrahmanyan, L., Bryant, Z., and Ruohola-Baker, H. (1997). maelstrom is required for an early step in the establishment of Drosophila oocyte polarity: posterior localization of grk mRNA. Development *124,* 4661-4671.

Cogoni, C., and Macino, G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. Nature *399,* 166-169.

Cook, H. A., Koppetsch, B. S., Wu, J., and Theurkauf, W. E. (2004). The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. Cell *116*, 817-829.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Costa, Y., Speed, R. M., Gautier, P., Semple, C. A., Maratou, K., Turner, J. M., and Cooke, H. J. (2006). Mouse MAELSTROM: the link between meiotic silencing of unsynapsed chromatin and microRNA pathway? Hum Mol Genet *15*, 2324-2334.

Covey, S. N., Al-Kaff, N. S., Langara, A., and Turner, D. S. (1997). Plants combat infection by gene silencing. Nature *385*, 781-782.

Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

Cox, D. N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development *127*, 503-514.

Cuomo, C. A., Güldener, U., Xu, J. R., Trail, F., Turgeon, B. G., Di Pietro, A., Walton, J. D., Ma, L. J., Baker, S. E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y. L., Decaprio, D., Gale, L. R., Gnerre, S., Goswami, R. S., Hammond-Kosack, K., Harris, L. J., Hilburn, K., Kennell, J. C., Kroken, S., Magnuson, J. K., Mannhaupt, G., Mauceli, E., Mewes, H. W., Mitterbauer, R., Muehlbauer, G., Münsterkötter, M., Nelson, D., O'donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M. I., Seong, K. Y., Tetko, I. V., Urban, M., Waalwijk, C., Ward, T. J., Yao, J., Birren, B. W., and Kistler, H. C. (2007). The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science *317*, 1400-1402.

Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J., and Brennecke, J. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Czech, B., Preall, J. B., McGinn, J., and Hannon, G. J. (2013). A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. Mol Cell

Dantonel, J. C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000). TBP-like factor is required for embryonic RNA polymerase II transcription in C. elegans. Mol Cell *6*, 715-722.

de Vanssay, A., Bouge, A. L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature *490*, 112-115.

Demerec, M. (1926a). Miniature-alpha-A Second Frequently Mutating Character in Drosophila Virilis. Proc Natl Acad Sci U S A *12*, 687-690.

Demerec, M. (1926b). Reddish-A Frequently "Mutating" Character in Drosophila Virilis. Proc Natl Acad Sci U S A *12*, 11-16.

Djupedal, I., Kos-Braun, I. C., Mosher, R. A., Söderholm, N., Simmer, F., Hardcastle, T. J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E., Wagner, E. G., Baulcombe, D. C., Allshire, R. C., and Ekwall, K. (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J *28*, 3832-3844.

Djupedal, I., Portoso, M., Spåhr, H., Bonilla, C., Gustafsson, C. M., Allshire, R. C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev *19*, 2301-2306.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Donertas, D., Sienski, G., and Brennecke, J. (2013). Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev *27*, 1693-1705.

Doolittle, W. F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Dorer, D. R., and Henikoff, S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell *77*, 993-1002.

Dufourt, J., Dennis, C., Boivin, A., Gueguen, N., Theron, E., Goriaux, C., Pouchin, P., Ronsseray, S., Brasset, E., and Vaury, C. (2013). Spatio-temporal requirements for transposable element piRNA-mediated silencing during Drosophila oogenesis. Nucleic Acids Res

Eddy, E. M. (1974). Fine structural observations on the form and distribution of nuage in germ cells of the rat. Anat Rec *178*, 731-757.

Eissenberg, J. C., and Elgin, S. C. (2000). The HP1 protein family: getting a grip on chromatin. Curr Opin Genet Dev *10*, 204-210.

Eissenberg, J. C., James, T. C., Foster-Hartnett, D. M., Hartnett, T., Ngan, V., and Elgin, S. C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc Natl Acad Sci U S A *87*, 9923-9927.

Eissenberg, J. C., Morris, G. D., Reuter, G., and Hartnett, T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. Genetics *131*, 345-352.

Fanti, L., Dorer, D. R., Berloco, M., Henikoff, S., and Pimpinelli, S. (1998). Heterochromatin protein 1 binds transgene arrays. Chromosoma *107*, 286-292.

Fedoroff, N. V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution. Science *338*, 758-767.

Findley, S. D. (2003). Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development *130*, 859-871.

Fu, Y., Wu, P., Beane, T., Zamore, P. D., and Weng, Z. (2018a). Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers.

Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K. A., Weng, Z., and Zamore, P. D. (2018b). The genome of the Hi5 germ cell line from. Elife *7*,

Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R. A. (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science *297*, 1871-1873.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep

Gowen, J. W., and Gay, E. H. (1934). Chromosome Constitution and Behavior in Eversporting and Mottling in Drosophila Melanogaster. Genetics *19*, 189-208.

Granados, R. R., Derksen, A. C., and Dwyer, K. G. (1986). Replication of the Trichoplusia ni granulosis and nuclear polyhedrosis viruses in cell cultures. Virology *152*, 472-476.

Grewal, S. I., and Elgin, S. C. (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr Opin Genet Dev *12*, 178-187.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., Degnan, B. M., Rokhsar, D. S., and Bartel, D. P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Grivna, S. T., Beyret, E., Wang, Z., and Lin, H. (2006a). A novel class of small RNAs in mouse spermatogenic cells. Genes Dev *20*, 1709-1714.

Grivna, S. T., Pyhtila, B., and Lin, H. (2006b). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc Natl Acad Sci U S A *103*, 13415-13420.

Gu, T., and Elgin, S. C. (2013). Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. PLoS Genet *9*, e1003780.

Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Ha, I., Lane, W. S., and Reinberg, D. (1991). Cloning of a human gene encoding the general transcription initiation factor IIB. Nature *352*, 689-695.

Haase, A. D., Fenoglio, S., Muerdter, F., Guzzardo, P. M., Czech, B., Pappin, D. J., Chen, C., Gordon, A., and Hannon, G. J. (2010). Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. Genes Dev *24*, 2499-2504.

Hall, I. M., Shankaranarayana, G. D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S. I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J *21*, 4671-4679.

Hamilton, A. J., and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Han, B. W., Wang, W., Li, C., Weng, Z., and Zamore, P. D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science *348*, 817-821.

Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2014). piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. Bioinformatics

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J *30*, 3977-3993.

Hansen, K. R., Burns, G., Mata, J., Volpe, T. A., Martienssen, R. A., Bähler, J., and Thon, G. (2005). Global effects on gene expression in fission yeast by silencing and RNA interference machineries. Mol Cell Biol *25*, 590-601.

Harr, J. C., Gonzalez-Sandoval, A., and Gasser, S. M. (2016). Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. EMBO Rep *17*, 139-155.

Hay, B., Ackerman, L., Barbel, S., Jan, L. Y., and Jan, Y. N. (1988). Identification of a component of Drosophila polar granules. Development *103*, 625-640.

Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S. L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. Nature *539*, 588-592.

Herr, A. J., Jensen, M. B., Dalmay, T., and Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Hirose, Y., and Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events. Genes Dev *14*, 1415-1429.

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P. D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Howe, F. S., Fischl, H., Murray, S. C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation. Bioessays *39*, 1-12.

Huang, X., Fejes Tóth, K., and Aravin, A. A. (2017). piRNA Biogenesis in Drosophila melanogaster. Trends Genet *33*, 882-894.

Hur, J. K., Luo, Y., Moon, S., Ninova, M., Marinov, G. K., Chung, Y. D., and Aravin, A. A. (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. Genes Dev *30*, 840-855.

Hutvagner, G., and Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol *9*, 22-32.

Imbalzano, A. N., Zaret, K. S., and Kingston, R. E. (1994). Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. J Biol Chem *269*, 8280-8286.

Ioannidis, J. P. (2005). Why most published research findings are false. PLoS Med *2*, e124.

Irvine, D. V., Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L., and Martienssen, R. A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. Science *313*, 1134-1137.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell *78*, 657-668.

Jacobs, S. A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jiao, X., Chang, J. H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. Mol Cell *50*, 104-115.

Jinek, M., and Doudna, J. A. (2008). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405-412.

Kaltenbach, L., Horner, M. A., Rothman, J. H., and Mango, S. E. (2000). The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during C. elegans embryogenesis. Mol Cell *6*, 705-713.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M., and Celniker, S. E. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol *3*, RESEARCH0084.

Kato, H., Goto, D. B., Martienssen, R. A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467-469.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T. N., Siomi, M. C., and Siomi, H. (2008). *Drosophila* endogenous small RNAs bind to Argonaute2 in somatic cells. Nature *453*, 793-797.

Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3′ end formation of PIWI-interacting RNAs in vitro. Mol Cell *43*, 1015-1022.

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., and Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Khurana, J. S., Wang, J., Xu, J., Koppetsch, B. S., Thomson, T. C., Nowosielska, A., Li, C., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell *147*, 1551-1563.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.

Kim, S. M., Dubey, D. D., and Huberman, J. A. (2003). Early-replicating heterochromatin. Genes Dev *17*, 330-335.

Klattenhoff, C., Bratu, D. P., McGinnis-Schultz, N., Koppetsch, B. S., Cook, H. A., and Theurkauf, W. E. (2007). Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Dev Cell *12*, 45-55.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., Zhang, F., Schultz, N., Koppetsch, B. S., Nowosielska, A., Seitz, H., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Klenov, M. S., Lavrov, S. A., Korbut, A. P., Stolyarenko, A. D., Yakushev, E. Y., Reuter, M., Pillai, R. S., and Gvozdev, V. A. (2014). Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res

Klenov, M. S., Sokolova, O. A., Yakushev, E. Y., Stolyarenko, A. D., Mikhaleva, E. A., Lavrov, S. A., and Gvozdev, V. A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc Natl Acad Sci U S A *108*, 18760-18765.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol *18*, 490-495.

Knight, S. W., and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Kopytova, D. V., Krasnov, A. N., Kopantceva, M. R., Nabirochkina, E. N., Nikolenko, J. V., Maksimenko, O., Kurshakova, M. M., Lebedeva, L. A., Yerokhin, M. M., Simonova, O. B., Korochkin, L. I., Tora, L., Georgiev, P. G., and Georgieva, S. G. (2006). Two isoforms of Drosophila TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. Mol Cell Biol *26*, 7492-7505.

Kowalczykowski, S. C., Lonberg, N., Newport, J. W., Paul, L. S., and von Hippel, P. H. (1980). On the thermodynamics and kinetics of the cooperative binding of bacteriophage T4-coded gene 32 (helix destabilizing) protein to nucleic acid lattices. Biophys J *32*, 403-418.

Kulkarni, M., Ozgur, S., and Stoecklin, G. (2010). On track with P-bodies. Biochem Soc Trans *38*, 242-251.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T. W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., and Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev *22*, 908-917.

Kutach, A. K., and Kadonaga, J. T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol *20*, 4754-4764.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lagarrigue, S., Hormozdiari, F., Martin, L. J., Lecerf, F., Hasin, Y., Rau, C., Hagopian, R., Xiao, Y., Yan, J., Drake, T. A., Ghazalpour, A., Eskin, E., and Lusis, A. J. (2013). Limited RNA editing in exons of mouse liver and adipose. Genetics *193*, 1107-1115.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lau, N. C., Robine, N., Martin, R., Chung, W. J., Niki, Y., Berezikov, E., and Lai, E. C. (2009). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. Genome Res *19*, 1776-1785.

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., and Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. A., and Toth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev *27*, 390-399.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y. C., Luo, Y., Sachidanandam, R., Toth, K. F., Patel, D., and Aravin, A. A. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev *28*, 1667-1680.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev *10*, 1683-1698.

Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., Sharma, P. P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D. H., Zamore, P. D., Miska, E. A., Sarkies, P., and Jiggins, F. M. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol *2*, 174-181.

Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., Kittler, E. L., Zapp, M. L., Klattenhoff, C., Schulz, N., Theurkauf, W. E., Weng,

Z., and Zamore, P. D. (2009a). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987-2993.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000, G. P. D. P. S. (2009b). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, X. Z., Roy, C. K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B. W., Xu, J., Moore, M. J., Schimenti, J. C., Weng, Z., and Zamore, P. D. (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell *50*, 67-81.

Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. Development *120*, 1201-1211.

Lieberman, P. M., and Berk, A. J. (1994). A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. Genes Dev *8*, 995-1006.

Lim, A. K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A *104*, 6714-6719.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol *11*, 576-577.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, Z. W., Shao, C. R., Zhang, C. J., Zhou, J. X., Zhang, S. W., Li, L., Chen, S., Huang, H. W., Cai, T., and He, X. J. (2014). The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet *10*, e1003948.

Llave, C., Kasschau, K. D., Rector, M. A., and Carrington, J. C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605-1619.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Ma, J. B., Ye, K., and Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., and Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., and Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J *24*, 800-812.

Martianov, I., Fimia, G. M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. Mol Cell *7*, 509-515.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev *19*, 1512-1517.

Matsumoto, N., Sato, K., Nishimasu, H., Namba, Y., Miyakubi, K., Dohmae, N., Ishitani, R., Siomi, H., Siomi, M. C., and Nureki, O. (2015). Crystal Structure and Activity of the Endoribonuclease Domain of the piRNA Pathway Factor Maelstrom. Cell Rep

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C., Gilmour, D. S., Albert, I., and Pugh, B. F. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

McCarty, D. M., DiRosario, J., Gulaid, K., Muenzer, J., and Fu, H. (2009). Mannitol-facilitated CNS entry of rAAV2 vector significantly delayed the neurological disease progression in MPS IIIB mice. Gene Ther *16*, 1340-1352.

McCLINTOCK, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McCLINTOCK, B. (1951). Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol *16*, 13-47.

Meehan, R. R., Kao, C. F., and Pennings, S. (2003). HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. EMBO J *22*, 3164-3174.

Mette, M. F., van der Winden, J., Matzke, M., and Matzke, A. J. (2002). Short RNAs can identify new candidate transposable element families in Arabidopsis. Plant Physiol *130*, 6-9.

Mevel-Ninio, M., Pelisson, A., Kinder, J., Campos, A. R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615-1624.

Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science *348*, 812-817.

Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. Cell *157*, 1364-1379.

Morlan, J. D., Qu, K., and Sinicropi, D. V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. PLoS One *7*, e42882.

Moshkovich, N., and Lei, E. P. (2010). HP1 recruitment in the absence of argonaute proteins in Drosophila. PLoS Genet *6*, e1000880.

Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Muerdter, F., Guzzardo, P. M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G. J. (2013). A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in Drosophila. Mol Cell

Muller, H. J. (1930). Types of visible variations induced by X-rays in Drosophila. J. Genet. *22*, 299-334.

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Nechaev, S., Fargo, D. C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.

Ni, J. Q., Zhou, R., Czech, B., Liu, L. P., Holderbaum, L., Yang-Zhou, D., Shim, H. S., Tao, R., Handler, D., Karpowicz, P., Binari, R., Booker, M., Brennecke, J., Perkins, L. A., Hannon, G. J., and Perrimon, N. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat Methods *8*, 405-407.

Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103-107.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S. I. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. Nat Genet *36*, 1174-1180.

Ohtani, H., Iwasaki, Y. W., Shibuya, A., Siomi, H., Siomi, M. C., and Saito, K. (2013). DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. Genes Dev *27*, 1656-1661.

Okamura, K., Chung, W. J., Ruby, J. G., Guo, H., Bartel, D. P., and Lai, E. C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.

Onodera, Y., Haag, J. R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613-622.

Orgel, L. E., and Crick, F. H. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604-607.

Ozcan, G., Ozpolat, B., Coleman, R. L., Sood, A. K., and Lopez-Berestein, G. (2015). Preclinical and clinical development of siRNA-based therapeutics. Adv Drug Deliv Rev *87*, 108-119.

Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (1997). Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. Cell *90*, 479-490.

Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol Cell *9*, 315-327.

Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Chikka, M. R., Bhadra, U., Birchler, J. A., and Elgin, S. C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.

Pandey, R. R., Homolka, D., Chen, K. M., Sachidanandam, R., Fauvarque, M. O., and Pillai, R. S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. PLoS Genet *13*, e1006956.

Pane, A., Jiang, P., Zhao, D. Y., Singh, M., and Schupbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J *30*, 4601-4615.

Pane, A., Wehr, K., and Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell *12*, 851-862.

Papai, G., Tripathi, M. K., Ruhlmann, C., Layer, J. H., Weil, P. A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature *465*, 956-960.

Parnell, T. J., Huff, J. T., and Cairns, B. R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. EMBO J *27*, 100-110.

Paro, R., and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc Natl Acad Sci U S A *88*, 263-267.

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am J Hum Genet *31*, 106-115.

Patzelt, E., Thalmann, E., Hartmuth, K., Blaas, D., and Kuechler, E. (1987). Assembly of pre-mRNA splicing complex is cap dependent. Nucleic Acids Res *15*, 1387-1399.

Pek, J. W., Lim, A. K., and Kai, T. (2009). Drosophila maelstrom ensures proper germline stem cell lineage differentiation by repressing microRNA-7. Dev Cell *17*, 417-424.

Pek, J. W., Ng, B. F., and Kai, T. (2012). Polo-mediated phosphorylation of Maelstrom regulates oocyte determination during oogenesis in Drosophila. Development *139*, 4505-4513.

Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. J Virol *81*, 1951-1960.

Pelisson, A., Song, S. U., Prud'homme, N., Smith, P. A., Bucheton, A., and Corces, V. G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J *13*, 4401-4411.

Peng, J. C., and Karpen, G. H. (2008). Epigenetic regulation of heterochromatic DNA stability. Curr Opin Genet Dev *18*, 204-211.

Petrie, V. J., Wuitschick, J. D., Givens, C. D., Kosinski, A. M., and Partridge, J. F. (2005). RNA interference (RNAi)-dependent and RNAi-independent association of the Chp1 chromodomain protein with distinct heterochromatic loci in fission yeast. Mol Cell Biol *25*, 2331-2346.

Post, C., Clark, J. P., Sytnikova, Y. A., Chirn, G. W., and Lau, N. C. (2014). The capacity of target silencing by Drosophila PIWI and piRNAs. RNA

Provost, P., Silverstein, R. A., Dishart, D., Walfridsson, J., Djupedal, I., Kniola, B., Wright, A., Samuelsson, B., Radmark, O., and Ekwall, K. (2002). Dicer is required for chromosome segregation and gene silencing in fission yeast cells. Proc Natl Acad Sci U S A *99*, 16648-16653.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697-711.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Quinlan, M. E. (2016). Cytoplasmic Streaming in the Drosophila Oocyte. Annu Rev Cell Dev Biol *32*, 173-195.

Rando, O. J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr Opin Genet Dev *22*, 148-155.

Rangan, P., Malone, C. D., Navarro, C., Newbold, S. P., Hayes, P. S., Sachidanandam, R., Hannon, G. J., and Lehmann, R. (2011). piRNA production requires heterochromatin formation in Drosophila. Curr Biol *21*, 1373-1379.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr Opin Cell Biol *17*, 269-273.

Reinhart, B. J., and Bartel, D. P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., and Pillai, R. S. (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat Struct Mol Biol *16*, 639-646.

Rhee, H. S., and Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Riddle, N. C., Minoda, A., Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Tolstorukov, M. Y., Gorchakov, A. A., Jaffe, J. D., Kennedy, C., Linder-Basso, D., Peach, S. E., Shanower, G., Zheng, H., Kuroda, M. I., Pirrotta, V., Park, P. J., Elgin, S. C., and Karpen, G. H. (2011). Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res *21*, 147-163.

Robert, V., Prud'homme, N., Kim, A., Bucheton, A., and Pélisson, A. (2001). Characterization of the flamenco region of the Drosophila melanogaster genome. Genetics *158*, 701-713.

Rogers, A. K., Situ, K., Perkins, E. M., and Toth, K. F. (2017). Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. Genes Dev *31*, 1858-1869.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Ronsseray, S., Anxolabehere, D., and Periquet, G. (1984). Hybrid dysgenesis in *Drosophila* melanogaster: influence of temperature on cytotype determination in the P-M system. Mol Gen Genet *196*, 17-23.

Rozhkov, N. V., Hammell, M., and Hannon, G. J. (2013). Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev *27*, 400-412.

Rubin, G. M., Kidwell, M. G., and Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell *29*, 987-994.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M. C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature *461*, 1296-1299.

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H., and Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev *24*, 2493-2498.

Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev *20*, 2214-2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M. C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics *166*, 1313-1321.

Sato, K., Nishida, K. M., Shibuya, A., Siomi, M. C., and Siomi, H. (2011). Maelstrom coordinates microtubule organization during Drosophila oogenesis through interaction with components of the MTOC. Genes Dev *25*, 2361-2373.

Schier, A. F. (2007). The maternal-zygotic transition: death and birth of RNAs. Science *316*, 406-407.

Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol *6*, 73-77.

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schotta, G., Ebert, A., Dorn, R., and Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. Semin Cell Dev Biol *14*, 67-75.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J *21*, 1121-1131.

Schrider, D. R., Gout, J. F., and Hahn, M. W. (2011). Very few RNA and DNA sequence differences in the human transcriptome. PLoS One *6*, e25842.

Schueler, M. G., and Sullivan, B. A. (2006). Structural and functional dynamics of human centromeric chromatin. Annu Rev Genomics Hum Genet *7*, 301-313.

Schultz, J. (1936). Variegation in Drosophila and the Inert Chromosome Regions. Proc Natl Acad Sci U S A *22*, 27-33.

Schupbach, T., and Wieschaus, E. (1986). Germline autonomy of maternal-effect mutations altering the embryonic body pattern of Drosophila. Dev Biol *113*, 443-448.

Schüpbach, T., and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. Genetics *121*, 101-117.

Schüpbach, T., and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics *129*, 1119-1136.

Senti, K. A., Jurczak, D., Sachidanandam, R., and Brennecke, J. (2015). piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. Genes Dev *29*, 1747-1762.

Shen, H. (2009). UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. BMB Rep *42*, 185-188.

Shpiz, S., and Kalmykova, A. (2014). Analyses of piRNA-mediated transcriptional transposon silencing in Drosophila: nuclear run-on assay on ovaries. Methods Mol Biol *1093*, 149-159.

Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M., and Kalmykova, A. (2011). Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. Nucleic Acids Res *39*, 8703-8711.

Sienski, G., Batki, J., Senti, K. A., Donertas, D., Tirian, L., Meixner, K., and Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev

Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964-980.

Sijen, T., and Plasterk, R. H. (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature *426*, 310-314.

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.

Smith, J. G., Caddle, M. S., Bulboaca, G. H., Wohlgemuth, J. G., Baum, M., Clarke, L., and Calos, M. P. (1995). Replication of centromere II of Schizosaccharomyces pombe. Mol Cell Biol *15*, 5165-5172.

Smothers, J. F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr Biol *10*, 27-30.

Smothers, J. F., and Henikoff, S. (2001). The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. Mol Cell Biol *21*, 2555-2569.

Song, J. J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Song, S. U., Kurkulos, M., Boeke, J. D., and Corces, V. G. (1997). Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of Drosophila. Development *124*, 2789-2798.

Soper, S. F., van der Heijden, G. W., Hardiman, T. C., Goodheart, M., Martin, S. L., de Boer, P., and Bortvin, A. (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev Cell *15*, 285-297.

Spofford, J. B. (1967). Single-locus modification of position-effect variegation in Drosophila melanogaster. I. White variegation. Genetics *57*, 751-766.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S. I. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proc Natl Acad Sci U S A *102*, 152-157.

Sun, F. L., Haynes, K., Simpson, C. L., Lee, S. D., Collins, L., Wuller, J., Eissenberg, J. C., and Elgin, S. C. (2004). cis-Acting determinants of heterochromatin formation on Drosophila melanogaster chromosome four. Mol Cell Biol *24*, 8210-8220.

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C. C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R. M., and Hannon, G. J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

Tartof, K. D., Hobbs, C., and Jones, M. (1984). A structural basis for variegating position effects. Cell *37*, 869-878.

Tenaillon, M. I., Hufford, M. B., Gaut, B. S., and Ross-Ibarra, J. (2011). Genome size and transposable element content as determined by high-throughput sequencing in maize and Zea luxurians. Genome Biol Evol *3*, 219-229.

Theurkauf, W. E., Smiley, S., Wong, M. L., and Alberts, B. M. (1992). Reorganization of the cytoskeleton during Drosophila oogenesis: implications for axis specification and intercellular transport. Development *115*, 923-936.

Tian, Y., Simanshu, D. K., Ma, J. B., and Patel, D. J. (2011). Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. Proc Natl Acad Sci U S A *108*, 903-910.

Todeschini, A. L., Teysset, L., Delmarre, V., and Ronsseray, S. (2010). The epigenetic trans-silencing effect in Drosophila involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. PLoS One *5*, e11032.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Travers, A. (2000). Recognition of distorted DNA structures by HMG domains. Curr Opin Struct Biol *10*, 102-109.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822-3831.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N., and Stuitje, A. R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell *2*, 291-299.

Veenstra, G. J., Weeks, D. L., and Wolffe, A. P. (2000). Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus. Science *290*, 2312-2315.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Vermaak, D., and Malik, H. S. (2009). Multiple roles for heterochromatin protein 1 genes in Drosophila. Annu Rev Genet *43*, 467-492.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Wang, H., Ma, Z., Niu, K., Xiao, Y., Wu, X., Pan, C., Zhao, Y., Wang, K., Zhang, Y., and Liu, N. (2016). Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in Drosophila. Development *143*, 530-539.

Wang, S. H., and Elgin, S. C. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci U S A *108*, 21164-21169.

Wang, W., Han, B. W., Tipping, C., Ge, D. T., Zhang, Z., Weng, Z., and Zamore, P. D. (2015). Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell *59*, 819-830.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M. A., Sakaki, Y., and Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., and Zamore, P. D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell *151*, 1055-1067.

Wehr, K., Swan, A., and Schüpbach, T. (2006). Deadlock, a novel protein of Drosophila, is required for germline maintenance, fusome morphogenesis and axial patterning in oogenesis and associates with centrosomes in the early embryo. Dev Biol *294*, 406-417.

Wickersheim, M. L., and Blumenstiel, J. P. (2013). Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. Biotechniques *55*, 269-272.

Wierzbicki, A. T., Ream, T. S., Haag, J. R., and Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.

Wilson, J. E., Connell, J. E., and Macdonald, P. M. (1996). aubergine enhances oskar translation in the Drosophila ovary. Development *122*, 1631-1639.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E., and Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol *2*, E104.

Xue, Y., Bai, X., Lee, I., Kallstrom, G., Ho, J., Brown, J., Stevens, A., and Johnson, A. W. (2000). Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. Mol Cell Biol *20*, 4006-4015.

Yang, N., and Kazazian, H. H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol *13*, 763-771.

Yao, B., La, L. B., Chen, Y. C., Chang, L. J., and Chan, E. K. (2012). Defining a new role of GW182 in maintaining miRNA stability. EMBO Rep *13*, 1102-1108.

Yoon, J., Lee, K. S., Park, J. S., Yu, K., Paik, S. G., and Kang, Y. K. (2008). dSETDB1 and SU(VAR)3-9 sequentially function during germline-stem cell differentiation in Drosophila melanogaster. PLoS One *3*, e2234.

Yu, B., Cassani, M., Wang, M., Liu, M., Ma, J., Li, G., Zhang, Z., and Huang, Y. (2015a). Structural insights into Rhino-mediated germline piRNA cluster formation. Cell Res *25*, 525-528.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J. B., Ma, J., Czech, B., and Hannon, G. J. (2015b). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science *350*, 339-342.

Zamore, P. D. (2017). Molecular biology: Rhino gives voice to silent chromatin. Nature *549*, 38-39.

Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T. F., Teng, C., Chen, X., Park, S. Y., Liu, L., Gallego-Bartolome, J., Liu, W., Henderson, I. R., Meyers, B. C., Ausin, I., and Jacobsen, S. E. (2015). A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. Cell *163*, 445-455.

Zhang, D., Penttila, T. L., Morris, P. L., Teichmann, M., and Roeder, R. G. (2001). Spermiogenesis deficiency in mice lacking the Trf2 gene. Science *292*, 1153-1155.

Zhang, D., Xiong, H., Shan, J., Xia, X., and Trudeau, V. L. (2008a). Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'-5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. Biol Direct *3*, 48.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B. S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871-884.

Zhang, H., Ma, Z. Y., Zeng, L., Tanaka, K., Zhang, C. J., Ma, J., Bai, G., Wang, P., Zhang, S. W., Liu, Z. W., Cai, T., Tang, K., Liu, R., Shi, X., He, X. J., and Zhu, J. K. (2013). DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A *110*, 8290-8295.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008b). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Z., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence *3*, 9.

Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S. S., Tu, S., Vreven, T., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2014). The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. Cell *157*, 1353-1363.

Zhang, Z., Xu, J., Koppetsch, B. S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W. E., and Zamore, P. D. (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell *44*, 572-584.

Zhao, X., and Herr, W. (2002). A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. Cell *108*, 615-627.

Zheng, C., and Hayes, J. J. (2003). Structures and interactions of the core histone tail domains. Biopolymers *68*, 539-546.

Zhong, X., Du, J., Hale, C. J., Gallego-Bartolome, J., Feng, S., Vashisht, A. A., Chory, J., Wohlschlegel, J. A., Patel, D. J., and Jacobsen, S. E. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell *157*, 1050-1060.